THE UNIVERSITY *of* EDINBURGH

Exploring biocatalysts for organic synthesis:

Mechanism, promiscuity and engineering

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Alexis James Antoine Hennessy

2020

Lay Abstract

As climate change and the protection of the environment are high on the global agenda, the chemical industry is adapting to this paradigm shift. Specifically, there is a necessity to find methodologies that avoid the creation of toxic waste and pollutants and the use of harsh conditions (high pressures or extreme temperatures). A fast-developing branch of chemistry, called green chemistry, supports the development of more environmentally-benign processes. An important aspect of green chemistry is biocatalysis. This consists in using biological systems, either whole cells or enzymes, for the production of industrially-relevant compounds. Indeed, biological systems, produce naturally a wide range of very complex chemical compounds under mild conditions (room temperature or 37°C in most cases, and atmospheric pressure), are very efficient and, in most cases, do not generate any toxic, unwanted side products. Moreover, enzymes are organic compounds, which means that they themselves can be recycled and do not contain any expensive and polluting heavy metals.

However, very often, enzymes will only be effective for their natural substrates, which are rarely interesting from a commercial point of view. To make them fit to produce other, more marketable products, they require to be further optimised and engineered. A suite of techniques and tools, called directed evolution, pioneered by 2018 Chemistry Nobel Prize winner, Prof. Frances Arnold (California Institute of Technology), has dramatically reduced the time and cost required to engineer enzymes. However, these methods require a basic understanding of their mechanism.

This work will focus on understanding the mechanism of two different transformations carried by specific enzymes. The first one, the Morita-Baylis-Hillman reaction, is a widely used reaction which makes it possible to couple together two molecules which have specific features, to create one larger, more complex compound. In the second one, the mechanism of a small family of enzymes called nitrile synthetases will be investigated, with the aim of engineering them to make amides, one of the most important type of molecules in the pharmaceutical industry as they represent more than 25% of all commercial drugs.

Abstract

In recent years, enzymes, and their uses as biocatalysts for synthetic organic chemistry, have been positively disrupting the pharmaceutical and chemical industries. With the development of the third wave of biocatalysis and directed evolution, enzyme engineering has become easier, cheaper and quicker. However, while random enzyme evolution is possible in the absence of structural data, the development of efficient biocatalysts is accelerated when 3D structures and mechanistic information of the target are available. In this work, two enzymes, each catalysing different chemical transformations will be studied, with the aim of gaining structural and mechanistic data which shall be used to guide future engineering strategies.

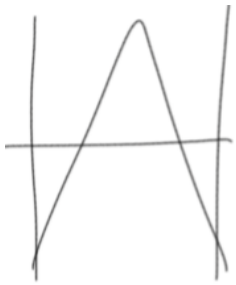
The first enzyme studied is the pimeloyl-ACP methyl ester esterase, BioH, from *E. coli* (EcBioH). Previously it was found that EcBioH is able to catalyse the Morita-Baylis-Hillman (MBH) coupling of 4-nitrobenzaldehyde (4-NBA) and methyl vinyl ketone (MVK), with less than 50% conversion. The MBH reaction forms a C-C bond by coupling an aldehyde and an electron-poor alkene, and leads to the formation of a chiral alcohol. However, the mechanism of the enzyme-catalysed biotransformation is unclear. Specifically the identities of the active residues involved in substrate binding and catalysis are not known. Here we describe the characterisation of recombinant EcBioH and study its role in catalysing the MBH reaction. We used site directed mutagenesis to probe the role of a proposed catalytic triad of conserved residues (Ser-Asp-His) in the mechanism. We also investigate the involvement of a cysteine residue by covalent modification and mass spectrometry. Since none of these residues were found to be essential for MBH catalysis, this suggested that the protein hexahistidine (also known as 6His) affinity tag used to purify the enzyme displays inherent MBH activity. This hypothesis was confirmed by screening other hexahistidine tagged proteins available in the laboratory and discovering that they also displayed MBH activity. Furthermore, an engineered EcBioH “double-tagged” at both the N- and C- termini displayed the highest MBH activity. No enzyme displayed enantioselectivity in the MBH product distribution. These results provide strong evidence that this commonly-used tag plays an important role in MBH biocatalysis. These observations suggested that the intrinsic MBH activity of the tag arises from a synergistic acid/base relationship between the histidine residues and oxo-anion intermediates.

In a second, independent project, a small family of enzymes, the nitrile synthetases (NS) was investigated. Interestingly, members of this family catalyse the conversion of a carboxylic acid to a nitrile in an ATP-dependent process. The reaction proceeds in two steps; the first involves the

formation of an amide intermediate, and the second step results in amide to nitrile conversion. There is still a lack of clarity of the mechanistic role of ATP in the two steps catalysed by NS. Most notably, no attempts at separating/decoupling the two steps catalysed by the NSs have been successful so far. In this work, two members of the NS family, QueC from *Bacillus subtilis* (BsQueC) and ToyM from *Streptomyces rimosus* (SrToyM) were purified from *E. coli* and characterised. Both NSs were assayed for their natural reaction, the transformation of 7-carboxydeazaguanine (CDG) into 7-cyanodeazaguanine (PreQ0), via the corresponding primary amide (ADG). Their substrate scope was assessed and it was found that both enzymes had a very restricted acid substrate selectivity but were able to form a small range of secondary and tertiary amides with primary and secondary amines. Crystal trials resulted in the preparation of diffraction-quality crystals and the determination of the X-ray structure of the BsQueC:ADP complex at a resolution of 2.1 Å. This structure, was combined with the structure of a previously published BsQueC:phosphate complex and a sequence alignment of various members of the NS family. This analysis located well conserved residues and guided rational engineering of BsQueC. Well-conserved hydrophilic residues in close proximity to the phosphate moiety of the ADP molecule identified three residues for mutagenesis. A large decrease of the amide to nitrile conversion activity was observed for two of the single-point mutants generated (BsQueC K163A and BsQueC R204A). Importantly, a double-mutant enzyme which combines both these mutations (BsQueC K163A R204A) was not able to catalyse nitrile formation. However, this rationally-designed biocatalyst displayed *bone-fide* amide synthetase activity and allowed kinetic parameters for the CDG substrate to be determined for the first time. Finally, the first large scale biocatalysed transformation of CDG to ADG was carried out and the amide product was isolated with >95% purity. These studies provide the necessary insight and strategy required to carry out further directed evolution of the NS family to expand their substrate scope and allow synthesis of commercially-relevant amide- and nitrile-containing compounds.

Declaration

I, Alexis James Antoine Hennessy, hereby declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree or professional qualification. Except where otherwise stated by reference or acknowledgment, the work presented is entirely my own.



Alexis J. A. Hennessy

Cambridge, September 2020.

Acknowledgements

First and foremost, I would like to thank Professor Dominic Campopiano for giving me the invaluable opportunity to join his research group and carry the interesting, yet challenging, work that will be presented in this thesis. I also thank Dom for his trust, his enthusiasm and for many inspiring discussions, whether scientific or non-scientific. I look forward to many other Celtic FC adventures in the future!

Thanks to Dr Jon Marles-Wright for the great help with the structural biology studies. Even at tough crystallographic times, his advice and positivity were key in obtaining the structure that will be presented in this work.

While the group I leave is very different from the group I joined, I have always felt privileged to be part of such a supportive and fun team. Working with great scientists with such a wide range of expertise in chemistry and biology and such diverse personalities was an amazing experience. I am grateful to all past and present members of the Campopiano group.

First, I am extremely grateful to the students I supervised: Wenli, Chloé, Josh and Stacey for their huge contributions on this work and for many insightful discussions.

Thanks to Annabel for the fun and the wisdom that she brought in the lab and into my life. Thank you for teaching me almost everything I know about molecular biology. I enjoyed our creativity in making great TikTok videos way before TikTok was cool and look forward to many more to come!

Thanks to Peter for teaching me so much about leadership. In the lab, I will always remember him as a tough but fair leader and it was challenging to maintain the lab up to his standards after he spread his wings. I look forward to many more adventures.

Thanks to Jo for her kindness and for the laughs. I thank her help in navigating the wild world of mass spectrometry and the less wild world of good Edinburgh coffee establishments.

Thanks to Silvia and Gary for being my travel companions in this long but rewarding PhD journey. Thanks to Silvia for her kindness, her positivity and for the fantastic food! Thanks to Gary for the massages!

Thanks to Piera for being a fantastic teacher to me at the beginning of this work and for her help at many stages along the way. Thank you for all the fun and smiles! I look forward to seeing you in Cambridge!

Thanks to Ben for forcing me to research more formal ways to thank him for bringing “good bants” into the lab. Thanks also to Ben for always having one screen on Grand Chelem tennis tournaments and for the consequent loss of productivity during these weeks.

Thanks to Catherine for our entertaining conversations on global affairs; to Shona for the laughs (I look forward to knowing “what’s up?”); to Chris for teaching me so many useful tricks; and to Mark for the best distractions.

I would also like to thank Amanda, Richard, Van, Alex, Bohdan, Micheal, Eva, Alice, Rhona, Zenam and all our project students and interns for their help and for making the lab a better place to work. I look forward to staying in touch with you.

Thanks also go to all the NMR, MS, Fermentation and Stores teams and to the rest of the School support services, without whom this work would have taken much longer. Thanks to Elaine for her never-failing support with regards to demonstrating.

Thanks to Vero for being almost an honorary member of the group. I am looking forward to many more fondue-fuelled adventures with Pete and Annabel and you, in Paris, Edinburgh or elsewhere.

Thanks to everyone in the School of Chemistry for creating such a fantastic atmosphere. A special thanks to Alan, Sandy and Minas for helping me keep memories of more carefree undergraduate times and to fellow Senator Brian for the great work we have or have not achieved in the *Senatus Academicus*.

Thanks to my team at Edinburgh Innovations for making it so easy for me to balance my PhD research with my commitments and for teaching me how great innovations can be commercialised efficiently. Thanks also to the Chemistry Society 2018/19 committee for a challenging but fantastic year!

A big thanks to my non-Chemistry PhD friends for bearing with my Chemistry-related problems. Special thanks go to Lilas, Leo, Rebecca and Marcin.

I would like to thank my parents and my family for their unwavering and never-failing support and encouragement at every step of a journey that started well before this PhD.

Finally, I would like to thank Chiara for bringing her fun, her smiles and her kindness into my life and for providing so many rays of sunshine through the clouds.

List of abbreviations

aaRS	Aminoacyl-tRNA synthetase
ABHF	α/β-hydrolase fold
ABS	Amide bond synthetase
ACP	Acyl carrier protein
ACS	American Chemical Society
ACS2	Acyl-coenzyme A synthetase
ADG	7-amidodeazaguanine
AdK	Adenylate kinase
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
AmS	Amide synthetase
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AUC	Area under the curve
B.	*Bacillus*
BIA	Bicyclic imidazolyl alcohol
BioF	8-amino-7-oxononanoate synthase
BioG	Pimeloyl-ACP methyl ester esterase
BioH	Pimeloyl-ACP methyl ester esterase
BioK	Pimeloyl-ACP methyl ester esterase
BioV	Pimeloyl-ACP methyl ester esterase
br s	Broad singlet
BSA	Bovine serum albumin
BsQueC	*B. subtilis* QueC
BsQueC-eAmS	Amide synthetase engineered from *B. subtilis* QueC
BsQueF	*B. subtilis* QueF
CAGR	Compound annual growth rate
CCP4	Collaborative Computational Project No. 4
CDG	7-carboxydeazaguanine
CoA	Coenzyme A
CP	Carrier protein
CPH_4	6-carboxy-5,6,7,8-tetrahydropterin
C-ter	C-terminal
CV	Column volume
d	Doublet
DABCO	1,4-diazabicyclo[2.2.2]octane
DCC	Dicyclohexyl carbodiimide
$dCDCl_3$	Deuterated chloroform
DCM	Dichloromethane
dDMSO	Deuterated dimethylsulfoxide
DERA	Desoxyribose 5-phosphate aldolase
DMAP	4-dimethyl aminopyridine

DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Desoxyribonucleic acid
D-PhgAT	D-phenylglycine aminotransferase
DTT	Dithiothreitol
E.	*Escherichia*
EcBioH	*E. coli* BioH
EcBioH-C	C-terminal histagged *E. coli* BioH
EcBioH-N	N-terminal histagged *E. coli* BioH
EcBioH-NC	N- and C-termini histagged *E. coli* BioH
EcBioH-x	Untagged *E. coli* BioH
ee	Enatiomeric excess
eq.	Equivalent
ESI	Electrospray ionisation
EtOAc	Ethyl acetate
EW	Electron-withdrawing
FabH	3-oxoacyl-ACP-synthase
FTICR-MS	Fourier-transform ion cyclotron resonance mass spectrometry
Goase	Galactose oxidase
GSK	GlaxoSmithKline
GTP	Guanosine triphosphate
H.	*Haemophilus*
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HEPES	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonate
HiBioG	*H.influenzae* BioG
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
HTP	High throughput
IAA	Iodoacetamide
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-B-D-thiogalactopyranoside
k_{cat}	Catalytic constant
KCl	Potassium chloride
k_M	Michaelis constant
LB	Lysogeny broth
LC	Liquid chromatography
LC-MS	Liquid chromatography - mass spectrometry
m	Multiplet
MBH	Morita-Baylis-Hillman
MeCN	Acetonitrile
MES	2-(N-morpholino)ethanesulfonic acid
MgATP	Magnesium adenosine triphosphate

MR	Molecular replacement
MS	Mass spectrometry
MVK	Methyl vinyl ketone
NaCl	Sodium chloride
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NEB	New England Biolabs
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthase
NS	Nitrile synthetase
N-ter	N-terminal
OD_{600}	Optical density at 600 nm
P.	*Pectobacterium*
PanK	Pantothenate kinase
PBS	Phosphate buffer saline
PCAS	Pimeloyl-coenzyme A synthetase
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
Pi	Phosphate
PI3Kδ	Phosphatidylinositol-3-OH kinase
PLP	Pyridoxal 5'-phosphate
PMS	Phenylmethylsulfonyl
PMSF	Phenylmethylsulfonyl fluoride
PNP	Purine nucleoside phosphorylase
pNP	*p*-nitrophenol
pNPA	*p*-nitrophenylacetate
PPM	Phosphopentomutase
PPT	Phosphotransferase
$PreQ_0$	7-cyanodeazaguanine
QueC	7-cyanodeazaguanine synthetase
QueF	7-cyanodeazaguanine reductase
RC	Rauhut-Currier
rpm	Rotations per minute
RT	Room temperature
Rt	Retention time
s	Singlet
S.	*Streptomyces*
(S)-BINAP	(S)-(−)-(1,1-Binaphthalene-2,2-diyl)bis(diphenylphosphine)
SAM	S-adenosyl methionine
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SeMet-SrToyM	Selenomethionine enriched *S. rimosus* ToyM
SOC	Super optimal broth

SrToyM	*S. rimosus* ToyM
TCA	Trichloroacetic acid
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
T_m	Melting temperature
TOTT	2-(1-Oxy-pyridin-2-yl)-1,1,3,3-tetramethylisothiouronium tetrafluoroborate
ToyM	7-cyanodeazaguanine synthetase
TPP	Thiamine pyrophosphate
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer ribonucleic acid
UV	Ultraviolet
WT	Wild type
4-NBA	4-nitrobenzaldehyde

Table of Contents

Chapter 1 Introduction 21

1.1 Background to biocatalysis 21

1.1.1 Biocatalysis in a nut shell 21

1.1.2 Limitations of Biocatalysis 24

1.1.3 “Biocatalomics”: the economic aspects of biocatalysis 24

1.1.4 Biocatalysis in the industrial production of fine chemicals 26

1.2 Examples of organic reactions in biocatalysis 27

1.2.1 Stetter reaction 27

1.2.2 Rauhut-Currier reaction 29

1.2.3 Morita-Baylis-Hillman (MBH) reaction 30

1.2.4 Amide coupling 31

Chapter 2 Towards a biocatalytic approach for the Morita-Baylis-Hillman Reaction 34

2.1 Introduction 34

2.1.1 The Morita-Baylis-Hillman (MBH) reaction 34

2.1.2 Literature review of biocatalysed MBH reactions 38

2.1.3 Alpha-beta hydrolase fold, an interesting structural motif 39

2.1.4 EcBioH, a promiscuous esterase 41

2.1.5 Aims 48

2.2 Generation of the biocatalyst 49

2.2.1: Cloning of *E. coli* BioH 49

2.2.2: Expression and purification of EcBioH 50

2.3 Development of robust assays 53

2.3.1 A robust, medium-throughput reverse phase HPLC MBH assay 54

2.3.2 Chiral HPLC MBH assay 55

2.3.2 Investigating alternative high throughput MBH assays 55

2.3.4 Esterase reaction assay 57

2.4 Optimisation of the biocatalysed MBH reaction 59
2.4.1 Enzyme loading 59
2.4.2 Temperature 61
2.4.3 Concentration of reagents 61
2.5 Investigating the involvement of the catalytic triad 64
2.5.1 Generation of the mutants 65
2.5.2 Assaying the mutants 66
2.5.3 Mass spectrometry analysis of the EcBioH covalent product 67
2.6 Investigating the involvement of cysteine 71
2.6.1 Cysteine, another potent nucleophile 71
2.6.2 Cysteine alkylation experiments 71
3.7 Investigating the effect of the folding of the enzyme 73
2.7.1 Investigating another pimeloyl-ACP methyl ester esterase 74
2.7.2 BioG, an atypically folded enzyme 74
2.7.3 Assaying HiBioG 76
2.8 Assaying the involvement of histidine in the EcBioH-catalysed MBH reaction 78
2.8.1 Histidine, a weak catalyst for the chemical MBH reaction 78
2.8.2 Proposed role of acid/base catalysis 80
2.8.3 Investigating the involvement of the hexahistidine tag 81
2.8.5 Assaying EcBioH variants with different number of hexahistidine tags 85
2.8.6 Further investigation of the hexahistidine tag in the MBH reaction 88
2.8.6.1 *E. coli* Acyl carrier protein (ACP) 88
2.8.6.2 *E. coli* 8-amino-7-oxononanoate synthase (BioF) 88
2.8.6.3 *E. coli* 3-oxoacyl-ACP-synthase (FabH) 89
2.8.6.4 D-phenylglycine aminotransferase (D-PhgAT) 90
2.8.6.5 Assaying these enzymes 91
2.9 A general residue catalysis? 92

2.10 Conclusion 92
References 94
Chapter 3 Overcoming evolution to expand the synthetic scope of a nitrile synthetase 97
3.1 Introduction 97
3.1.1 Nitrile synthetase, an interesting family of enzymes 97
3.1.2 NS in natural products biosynthesis 98
3.1.3 Overview of QueC and ToyM 102
3.1.4 Chemical approaches to amide synthesis from carboxylic acid 106
3.1.4.1 Coupling reagents 107
3.1.4.2 Making amides via the acyl chloride 108
3.1.5 Amide bond formation using biocatalysis 110
3.1.5.1 ATP-dependent amide synthetases 110
3.1.5.2 The biocatalytic approach in action: examples of industrial uses 112
3.1.6 Making nitriles from carboxylic acids. 115
3.1.6.1 The direct route 115
3.1.6.2 The amide route 116
3.1.7 Aims 117
3.2 Generation of the biocatalysts 117
3.2.1 Generation of BsQueC 118
3.2.2 Generation of *SrToyM* 120
3.3 Generation of the substrates and products 123
3.3.1 Syntheses of CDG and $PreQ_0$ 123
3.3.2 Synthesis of ADG 125
3.4 Towards a nitrile synthetase assay 127
3.4.1 Development of an HPLC assay 127
3.4.2 Can an efficient HT coupled assay be found ? 128
3.5 Assaying NSs 131

3.5.1 Optimisation of reaction conditions 131
3.5.1.1 Enzyme loading 132
3.5.1.2 MgATP equivalents 133
3.5.1.3 CDG loading 135
3.6 Substrate screens 138
3.6.1 Carboxylic acid substrate screen 138
3.6.2 Amine screen 141
3.7 Structural studies 144
3.7.1 BsQueC 144
3.7.1.1 HT conditions screens on the unbound enzyme 144
3.7.1.2 Optimisation 145
3.7.1.3 BsQueC + ADP structure 148
3.7.2 SrToyM 152
3.7.2.1 HT conditions screens on the ligand-free enzyme 152
3.7.2.2 Generation of Selenomethionine-enriched SrToyM 153
3.7.2.3 Optimisation of Conditions A 154
3.7.2.4 Optimisation of Conditions B 158
3.7.2.5 Optimisation of Conditions C 158
3.7.2.6 SrToyM + CDG co-crystals 160
3.7.2.7 Structure of SrToyM 161
3.8 Engineering an amide synthetase 161
3.8.1 Identifying conserved residues 162
3.8.2 Finding conserved residues that are not involved in MgATP or Zn^{2+} binding 163
3.8.3 Finding conserved residues in the vicinity of the ADP 164
3.8.4 Generation and assay of the BsQueC mutants 165
3.8.5 Generation of the double-mutant 169
3.9 Engineering an amide synthetase from SrToyM 172

3.10 Characterisation of BsQueC-eAmS 174
3.10.1 BsQueC-eAmS kinetic studies 174
3.10.2 BsQueC-eAmS structural study 177
3.10.3 ADG synthesis using the BsQueC-eAmS biocatalyst 178
3.11 Conclusion 182
References 184
Chapter 4 Summary and outlook 187
Chapter 5 Materials and Methods 190
5.1 Materials 190
5.1.1 Growth media 190
5.1.2 Competent cell lines 190
5.1.3 Plasmids 191
5.1.4 Buffers 191
5.2 Molecular cloning 192
5.2.1 Overview 192
5.2.2 General procedures 192
5.2.2.1 Gel electrophoresis and extraction 192
5.2.2.2 Restriction digest 193
5.2.2.3 Analytical digest 193
5.2.2.4 Ligation 193
5.2.2.5 Miniprep 193
5.2.2.6 Site-directed mutagenesis 193
5.2.3 EcBioH initial cloning 195
5.2.3.1) Amplification and purification of ECBioH gene 195
5.2.3.2) pGEM ligation and analysis 196
5.2.3.3) Vector transfer of the insert: from pGEM-T vector to pET28a vector 196
5.3 Protein production 197

5.3.1 Transformation 197
5.3.2 Expression tests 197
5.3.3 Large scale expression 198
5.3.4 Expression of SeMet-SrToyM 198
5.3.5 Cell harvesting 199
5.4 Protein purification 199
5.4.1 Cell lysis 199
5.4.2 IMAC purification 199
5.4.3 Anion exchange purification of untagged EcBioH 200
5.4.4 Size-exclusion chromatography 200
5.5 Protein characterisation 201
5.5.1 Determination of enzyme concentration using Bradford assay 201
5.5.2 SDS-PAGE 202
5.5.3 Protein mass spectrometry 202
5.6 Organic synthesis 202
5.6.1 Model MBH reaction 202
5.6.2 Organic synthesis of Chapter 4 compounds 203
5.6.2.1. Synthetic route to CDG and $PreQ_0$ 203
5.6.2.2 One-pot synthesis of CDG-methyl ester 203
5.6.2.3 One-pot synthesis of $PreQ_0$ 204
5.6.2.4 Synthesis of CDG 205
5.7 Protein reactions for Chapter 3 205
5.7.1 Biocatalysed MBH reaction optimised protocol 205
5.7.2 Biocatalysed MBH reaction HPLC assay 205
5.7.3 Biocatalysed MBH reaction chiral HPLC assay 206
5.7.4 pNPA esterase assay 206
5.7.5 MS assay 206

5.7.5.1 PMSF 206
5.7.5.2 IAA 206
5.7.6 Buffer exchange 206
5.8 Protein reactions for Chapter 4 207
5.8.1 NS reaction optimised protocol 207
5.8.2 Large scale biosynthesis of ADG using BsQueC-eAmS 207
5.8.3 HPLC assay 208
5.8.4 BsQueF reductase assay 208
5.8.5 BsQueC-eAmS kinetics study 208
5.8.6 X-ray crystallography 208
5.8.6.1 Robot screen 208
5.8.6.2 Optimisation protocol 209
5.8.6.3 ADP-bound BsQueC structure 209
Appendix 1: EcBioH vector maps 210
Appendix 2: *E. coli bioH* gene sequence 212
Appendix 3: EcBioH-N protein sequence (the hexahistidine tag is in bold) 212
Appendix 4: EcBioH-C protein sequence (the hexahistidine tag is in bold) 212
Appendix 5: EcBioH-x protein sequence 212
Appendix 6: EcBioH-NC protein sequence (the hexahistidine tags are in bold) 212
Appendix 7: *H.influenzae bioG* gene sequence 213
Appendix 8: *H.influenzae* BioG protein sequence 213
Appendix 9: Calibration curve of the Superdex HiLoad 16/60 S200 gel filtration column (taken from Dr Piera Marchetti PhD thesis) 214
Appendix 10: Calibration curve of the Superdex HiLoad 16/60 S75 gel filtration column (taken from Dr Piera Marchetti PhD thesis) 215
Appendix 11: HPLC calibration curves for 4-NBA and the MBH product 216
Appendix 12: ^{1}H NMR of the MBH product 217

Appendix 13: *B. subtilis* QueC gene sequence 218

Appendix 14: BsQueC protein sequence (the hexahistidine tag is in bold) 218

Appendix 15: *B. subtilis* QueF gene sequence 218

Appendix 16: BsQueF protein sequence (the hexahistidine tag is in bold) 218

Appendix 17: *S. rimosus* ToyM gene sequence 219

Appendix 18: SrToyM protein sequence (the hexahistidine tag is in bold) 219

Appendix 19: NS vector maps 220

Appendix 20: HPLC calibration curves for CDG and $PreQ_0$ 222

Appendix 21: 7-carboxydeazaguanine ^{1}H NMR 223

Appendix 22: 7-carboxydeazaguanine ^{13}C NMR 223

Appendix 23: 7-cyanodeazaguanine ^{1}H NMR 224

Appendix 24: 7-cyanodeazaguanine ^{13}C NMR 224

Appendix 25: BsQueC crystallography data refinement and refinement statistics 225

Chapter 1 Introduction

1.1 Background to biocatalysis

1.1.1 Biocatalysis in a nut shell

The world's oldest brewery was found in a prehistoric burial site in Haifa, Israel[1]. There, traces of beer dating from 11 000 BC were found[1]. Knowing that yeast is required for the alcoholic fermentation of cereals and that agriculture was only developed around 10 000 BC[2], one may say that humanity was able to domesticate microorganisms and enzymes before plants or animals. Today, humans are able to harness the huge potential that enzymes offer to carry out a wide range of chemistries and produce a large array of complex compounds[3-6].

The study, engineering and optimisation of enzymes for applications in industrial synthetic chemistry defines the field of biocatalysis. This field has considerably developed, especially over the last few decades, culminating in the award of the 2018 Nobel Prize in Chemistry to Frances Arnold (Caltech), one of the pioneers of the field[7]. According to Uwe Bornscheuer and colleagues, this development can be divided in three parts, named the "three waves of biocatalysis"[8]. In the first wave, which started in the beginning of the 20th century, it was recognised that some components the living cells of animals, plants or microorganisms could be useful to promote useful chemical transformations such as the hydroxylation of steroids[9]. In the second wave of biocatalysis, which started in the 1980s with the advances of structural biology and the successful determination of high resolution protein structures, initial protein engineering approaches were introduced. These advances allowed the manufacture of pharmaceutical intermediates and fine chemicals, such as the calcium channel blocker diltiazem[10]. Finally, in the third wave of biocatalysis, which is still currently running, advances in molecular biology and bioinformatics enable the "directed evolution" of enzymes. This process moves the Darwinian process of evolution and natural selection into the laboratory as it consists of subjecting a gene to iterative rounds of mutagenesis, screening and selection to identify the variants which exhibit an improvement compared to the parent.

Biocatalysis is developing fast because it is a viable response to the global challenges attached to the high economic and environmental costs of chemical manufacture. Specifically, while homo- and heterogeneous catalysis often rely on energy-intensive processes and on the use of expansive second and third row transition metals, enzymes are biodegradable and require mild reaction conditions. Enzyme reactions are typically carried at temperatures ranging from room temperature (RT) to 50°C, at neutral pHs and at atmospheric pressure. In addition enzymes commonly exhibit desirable regio-,

diastero- and enantioselectivities which shall reduce the number of synthetic steps required to generate a given product and the need of protection/deprotection steps. Paul Anastas and John Warner have developed “The 12 principles of Green Chemistry”[11]. Table 1.1 displays these principles, with an emphasis on those for which biocatalysis can be a solution and how.

No	Principle	Biocatalysis solution
1	**Waste prevention** It is better to prevent waste than to treat or clean up waste after it has been created.	
2	**Atom economy** Synthetic methods should be designed to maximize incorporation of all materials used in the process into the final product.	
3	**Less hazardous chemical syntheses:** Wherever practicable, synthetic methods should be designed to use and generate substances that possess little or no toxicity to human health and the environment	Unlike most chemical catalysts, enzymes are biocompatible and do not exhibit any toxicity for humans or the environment. In addition, biocatalysis can enable the use of alternative starting materials and reagents with lower toxicity
4	**Designing safer chemicals** Chemical products should be designed to preserve efficacy of function while reducing toxicity.	
5	**Safer solvents and auxiliaries** The use of auxiliary substances (e.g., solvents, separation agents, etc.) should be made unnecessary wherever possible and, innocuous when used.	While a small percentage of organic solvents might be needed to put the reagents in solution, enzymes are active in water
6	**Design for energy efficiency** Energy requirements should be recognized for their environmental and economic impacts and should be minimized. Synthetic methods should be conducted at ambient temperature and pressure.	Enzyme reactions are typically carried at temperatures ranging from RT to 50°C and at atmospheric pressure. In addition, unlike most traditional catalysts, most enzymes are not air- or water-sensitive and do not require air or water elimination systems. Hence, enzymatic processes are typically highly energy efficient
7	**Use of renewable feedstocks** A raw material or feedstock should be renewable rather than depleting whenever technically and economically practicable.	Biocatalysts are fully biodegradable
8	**Reduce derivatives** Unnecessary derivatization (use of blocking groups, protection/deprotection, temporary modification of physical/chemical processes) should be minimized or avoided if possible, because such steps require additional reagents and can generate waste.	Enzymes commonly exhibit desirable regio-, diastero- and enantioselectivities which shall reduce the number of synthetic steps required to generate a given product and the need of protection/deprotection steps
9	**Catalysis** Catalytic reagents (as selective as possible) are superior to stoichiometric reagents.	Enzymes are catalysts.
10	**Design for degradation** Chemical products should be designed so that at the end of their function they break down into innocuous degradation products and do not persist in the environment.	As for Principle 7
11	**Real-time analysis for pollution prevention** Analytical methodologies need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances.	
12	**Inherently safer chemistry for accident prevention** Substances and the form of a substance used in a chemical process should be chosen to minimize the potential for chemical accidents, including releases, explosions, and fires.	As for Principles 3, 5 and 6, enzymes commonly require mild conditions and non-toxic solvents, reducing the risks of chemical accidents

Table 1.1: Solutions for the "12 principles of Green Chemistry" provided by biocatalysis. The principles and their definitions were taken from the American Chemical Society[12].

1.1.2 Limitations of Biocatalysis

Despite its previously discussed advantages, several drawbacks are associated with biocatalysis. The first one is the limited stability of typical biocatalysts which restrict the range of conditions (temperature, solvent, concentration of reagents…) that most biocatalysts can be used in and the possibilities of coupling chemical and biocatalytic processes. In addition, while biocatalysts can exhibit desirable regioselectivity, an associated downside is that most enzymes exhibit poor substrate scopes. For this reason, there is a very limited number of commercial, ready-to-use efficient enzymes, such as lipases. In the vast majority of cases, enzymes which carry the desired transformation will need to be screened for the given substrate and hits will necessitate further optimisation to have activities that are comparable with chemical catalysts. Currently, an increasing number of industrial biotechnology companies (Codexis, Prozomix, Biocatalysts, Johnson Matthey, Almac…) supply enzyme screening kits to assist users in finding enzymes that display activity to their specific substrate. In addition, as biocatalysis is poised to play an ever-increasing role in industrial chemistry, substantial research efforts are dedicated to developing tools to ease the development of novel biocatalysts. Especially, the development of novel bioinformatic tools should become increasingly efficient at predicting hot spots for mutations, which shall reduce the time required to develop novel biocatalysts.

1.1.3 "Biocatalomics": the economic aspects of biocatalysis

Unsurprisingly, biocatalysis is an expanding market. While the global enzyme market value was estimated to be $7 billion, it is estimated to reach $10 billion by 2024, with an impressive compound annual growth rate (CAGR) of 7.4%[13]. Biocatalysts have applications in a wide range of sectors, such as food processing, textile, pharma or biofuels and both for bulk and fine chemicals. While examples of fine chemicals produced partially or wholly using enzymes will be further discussed in the next section, an example of bulk chemical enzymatically produced is acrylamide. Acrylamide is produced by Nitto Chemical Company at a scale of 10 000+ tonnes per year using a nitrile hydratase enzyme, as shown in Figure 1.1, completely supplanting the previously used chemical approach that required a copper catalyst[14].

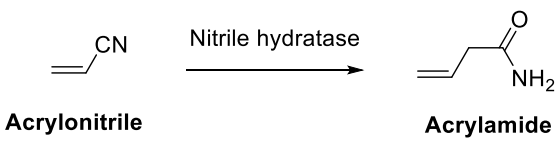


Figure 1.1: Example of bulk chemical produced by biocatalysis.

Examples of 20 compounds manufactured by biotransformation at scales higher than 100 tonnes per year, along with their applications and the type of enzyme used are shown on Table 1.2[15].

Production scale (tonnes per annum)	Product	Application(s)	Enzyme type
1000000+	High corn syrup	Food & beverage	Glucose isomerase
100000+	Lactose-free milk	Food & beverage	Lactase
10000+	Acrylamide	Biotechnology; contact lenses	Nitrile hydratase
	Cocoa butter	Food & beverage; cosmetics	Lipase
1000+	Nicotinamide	Dietary supplement	Itrilase
	D-pantothenic acid	Dietary supplement	Aldonolactonase
	(S)-chloropropinoic acid	Chemical building block	lipase
	6-aminopenillanic acid	Pharmaceutical (precursor)	Penicillin amidase
	Aspartame	Food & beverage	Thermolysin
	L-aspartate	Pharmaceutical; dietary supplement	Aspartase
	D-phenylglycine	Pharmaceutical (precursor)	Hydantoinase
	4-hydroxy-D-phenylglycine	Pharmaceutical (precursor)	Hydantoinase
100+	Ampicillin	Pharmaceutical	Penicillin amidase
	L-methionine, L-valine	Dietary supplement; chemical building block	Aminoacylase
	L-carnitine	Dietary supplement	Dehydratase/hydroxylase
	L-DOPA	Pharmaceutical	B-tyrosinase
	L-malic acid	Food & beverage; chemical building block	Fumarase
	(S)-methoxyisopropylamine	Chemistry	Lipase
	(R)-mandelic acid	Pharmaceutical (precursor)	Nitrilase
	L-alanine	Food & beverage; chemical building block	L-Asp-β-decarboxylase

Table 1.2: Range of applications of various large-scale biosynthesised products (Adapted from [15])

Table 1.2 highlights the wide range of industrial applications of biocatalysis and the large array of enzyme types that are used in industry. Majorly, biocatalysis has positively disrupted the industrial production of fine chemicals and some examples of this will be discussed in the next section.

1.1.4 Biocatalysis in the industrial production of fine chemicals

Many examples of enzymes are currently used in the pharmaceutical and fine chemical industries[16]. In this part, a few key examples will be cited. The use of a ketoreductase to induce chirality in the synthesis of Lipitor[17], a statin drug which was the second most prescribed drug in the United States in 2017[17]. Ketoreductases are, by far, the most used enzyme family in industry since, already in 2012, this family of enzymes was used to generate chiral alcohols in the route of 17 prescribed drugs[8]. Another enzyme family which has a high industrial potential is the transminases, which generate chiral amines. Notably, a highly engineered transaminase, which underwent 27 rounds of mutations to increase its activity and stability, was found to generate sitagliptin (Januvia)[18]. Another example of enzyme used in industry is a penicillin G acylase to produce amoxicillin[19]. Pathway engineering can also be achieved to "program" whole cells biosynthesis of active pharmaceutical ingredients. For example, Sanofi was able to engineer a yeast to produce the anti-malarial drug, Artemisinin[20]. Last but not least, an enzymatic cascade can be engineered to produce key molecules such as Islatravir[21]. The latter example will be further presented in the next section. All the examples are summarised in Table 1.3.

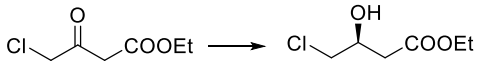
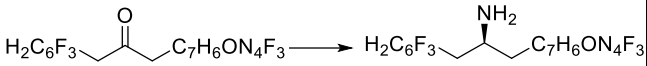
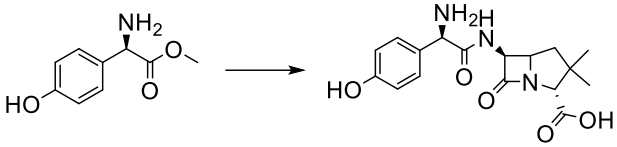
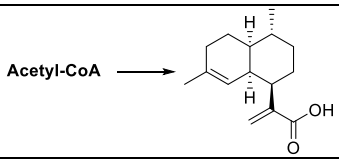
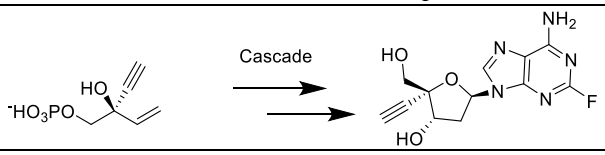
Company	Molecule	Enzyme family	Transformation	Revenues (billion $) or clinical phase
kaneka / CODEXIS PROTEIN ENGINEERING EXPERTS	**Lipitor**	Ketoreductase		1.5 (2018)[22]
MERCK INVENTING FOR LIFE / CODEXIS PROTEIN ENGINEERING EXPERTS	**Januvia**	Transaminase		4.5 (2020 projection)[23]
gsk	**Amoxicillin**	Penicillin G acylase		4.9 (2019)[24]
SANOFI	**Artemisinin**	Engineered yeast		0.7 (2025 projection)[25]
MERCK INVENTING FOR LIFE / CODEXIS PROTEIN ENGINEERING EXPERTS	**Islatravir**	Multiple		Phase III

Table 1.3: Examples of biocatalytic strategies used in the fine chemical industry.

1.2 Examples of organic reactions in biocatalysis

Name reactions are a set of chemical transformations which carry significant importance in organic synthesis. Many of these reactions have been observed in nature, in the biosyntheses of natural products and substantial research in the biocatalysis field have been dedicated at understanding the mechanisms of these enzymes and how they compare with chemical methods. Enzymes catalysing most classes of reactions such as condensation, rearrangement, oxidation and, even, halogenation were characterised[26]. Two examples of Name condensation reactions which were found to occur in biological systems, the Stetter reaction and the Rauhut-Currier reaction will be briefly discussed. Two other reactions where biocatalysis could potentially play a major role, the MBH reaction and the amide coupling reaction will also be discussed in this part.

1.2.1 Stetter reaction

The Stetter reaction is a condensation reaction between an α-β unsaturated ketone and an aldehyde, to form a dicarbonyl product, as shown on Figure 1.2. This reaction is traditionally catalysed by a cyanide ion. However, in nature, cofactor thiamine pyrophosphate (TPP) is used by enzymes to carry this transformation. For example, MenD, an enzyme in the biosynthetic pathway of menaquinone (vitamin K_2) catalyses the TTP-dependant Stetter reaction between 2-ketogluterate and

isochorismate, to form a 1,4-dicarbonyl[27]. The proposed mechanism of this reaction, shown in Figure 1.2, is thought to consist of two steps: in a first step, 2-ketoglutarate is decarboxylated to form a succinic semialdehyde-TPP anion. This anion will then carry nucleophilic attack on isochorismate, to form the product and regenerate the TPP cofactor.

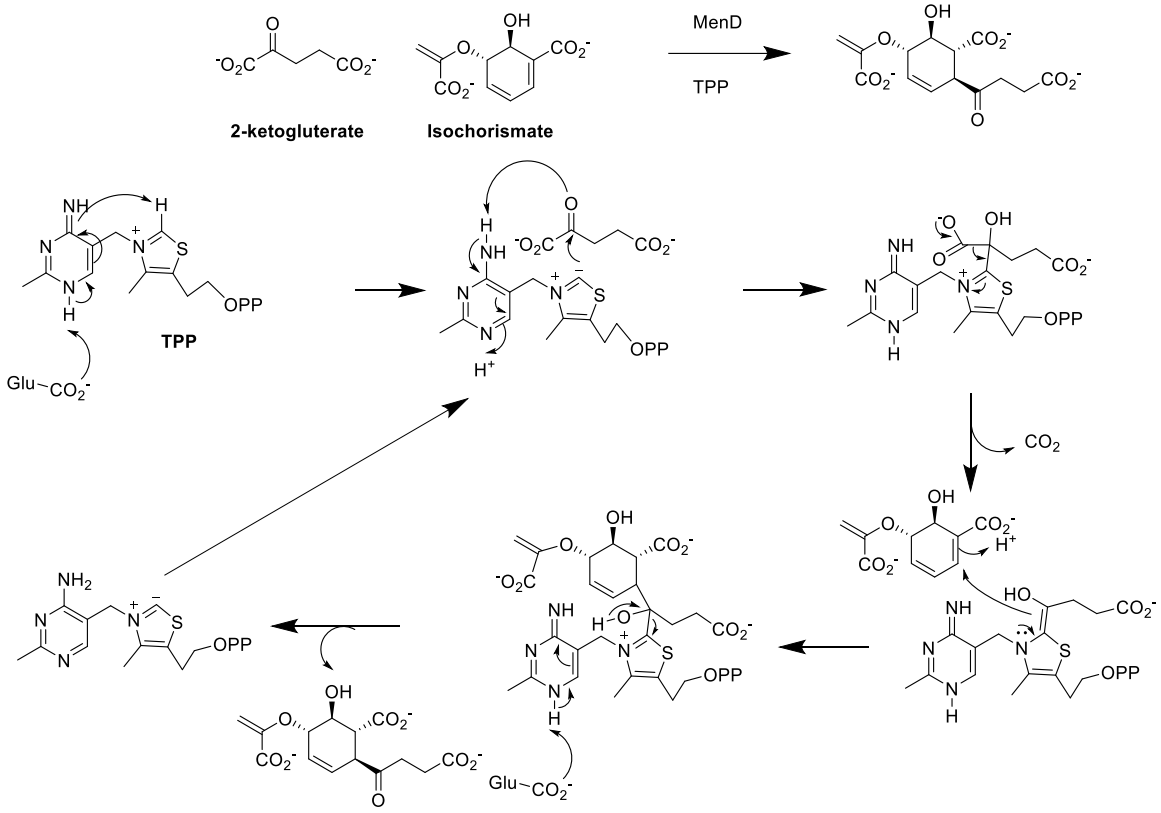


Figure 1.2: Biocatalysed MenD Stetter reaction scheme and mechanism: the reaction proceeds in two steps: first, 2-ketogluterate, bound to TTP, is decarboxylated to form an zwitterionic derivative which will carry nucleophilic attack on isochorismate. The adduct will then break down, generating the product.

Another example of biological Stetter reaction is catalysed by PigD from *Serratia marcescens* and involves the decarboxylative condensation of (2*E*)-octenal with pyruvate in the antiobiotic prodigiosin biosynthetic pathway, as shown in Figure 1.3[28].

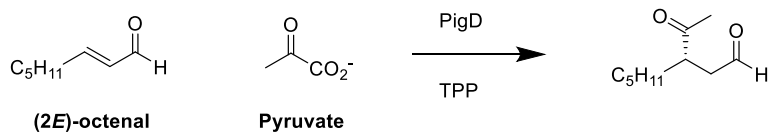


Figure 1.3: PigD reaction scheme.

1.2.2 Rauhut-Currier reaction

The Rauhut-Currier (RC) reaction is C-C bond-forming coupling twos electron-deficient alkenes. This reaction, shown in Figure 1.4, is often associated to the Morita-Baylis-Hillman (MBH) reaction, which will be extensively discussed in this work, in the sense that the RC reaction is thought to proceed *via* a similar mechanism. The main difference between the MBH and RC reactions is that, while the MBH reaction is the coupling of an electron-deficient alkene to an aldehyde, the RC reaction is the coupling of any two electron-withdrawing alkene[29]. Because of this difference, the RC reaction is often referred as a "vinylogous MBH" reaction. Interestingly, as it will be developed during this work, no enzyme catalysing a natural MBH reaction have been characterised. However, a few enzymes catalysing the RC reaction were found. An example of such enzymes is SpnL from *Saccharopolyspora spinosa*, an enzyme involved in the biosyntheses of spinosyns, a class of polyketide insecticides[30]. SpnL catalyses an intramolecular RC cyclisation. One of SpnL nucleophilic residues will undertake a nucleophilic attack on the substrate[30]. Interestingly, the identity of the nucleophile is unknown[30]. This will increase the nucleophilicity of the alkene and trigger the formation of a novel C-C bond. Rearrangement and regeneration of the catalyst will then afford the product. Its mechanism is shown in Figure 1.4.

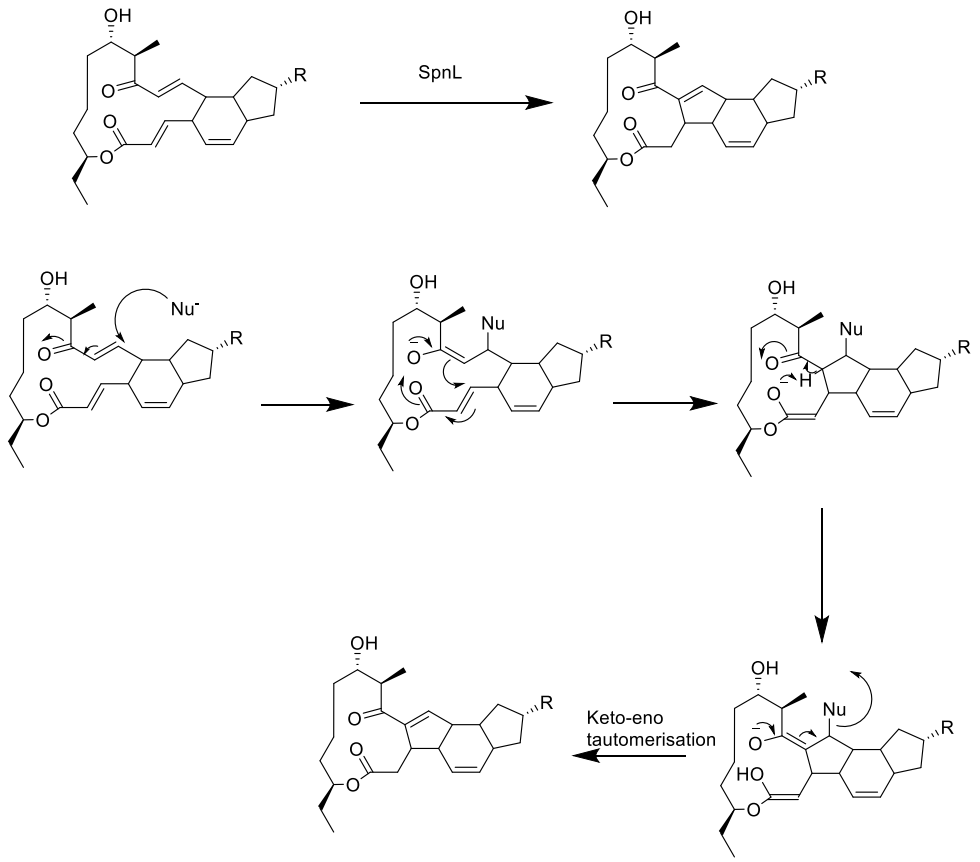


Figure 1.4: SpnL-catalysed RC reaction: the nucleophilic catalyst undertake Micheal addition on the substrate, generating a novel nucleophilic site which will form an intramolecular C-C bond, hence forming a novel 5-membered ring. Rearrangement and regeneration of the catalyst will afford the product.

1.2.3 Morita-Baylis-Hillman (MBH) reaction

As previously introduced, the MBH reaction is a more general variant of the RC reaction and represents the base-catalysed coupling of an electron-deficient alkene to an aldehyde. This reaction was extensively studied, as it will be discussed later in this work, and many organocatalysts have been developed[31]. However, no example of a natural "MBHase" enzyme have been reported in the literature and attempts at producing novel biocatalysts were based on exploiting the synthetic promiscuity of enzymes catalysing other reactions, such as esterases[32], lipases[33], halogenases[34] or racemases[34]. Amongst these attempts at finding a potent MBH biocatalyst, the pimeloyl-ACP methyl ester esterase BioH from *E. coli,* an esterase involved in the biotin biosynthetic pathway in numerous species was found to catalyse the coupling of 4-nitrobenzaldehyde and methyl vinyl ketone with a 46% conversion, as shown on Figure 1.5.

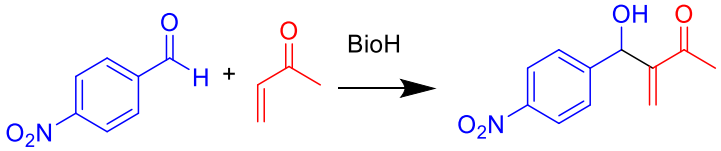


Figure 1.5: Scheme of the E.coli BioH-catalysed MBH reaction between 4-nitrobenzaldehyde and methyl vinyl ketone.

This represented a good starting point in the research of a potent MBH biocatalyst. As no mechanistic data was available, one of the aim of this work is to gain an understanding of the mechanism of this biocatalysed reaction through investigating the involvement of some features that could play a role in the reaction.

1.2.4 Amide coupling

By many metrics, the coupling of a carboxylic acid with an amine to generate an amide is amongst the most important reactions in synthetic organic chemistry. Chemically, this transformation requires the use of stoichiometric amounts of coupling reagents. For this reason, this reaction is considered atom inefficient and biocatalysis is thought to be a potential solution. Indeed, in nature many enzymes catalysing this reaction have been characterised. A comprehensive review of these will be carried later. Amongst these enzymes, an interesting small family of ATP-dependent enzymes have been described recently: the nitrile synthetases. These enzymes are capable of carrying the standard coupling of a carboxylic acid, 7-carboxydeazaguanine (CDG) and ammonium sulfate to generate the corresponding primary amide (7-amidodeazaguanine, ADG). However, the singularity of this enzyme family resides in that the amide is then further transformed into the corresponding nitrile ($PreQ_0$) as shown on Figure 1.6[36].

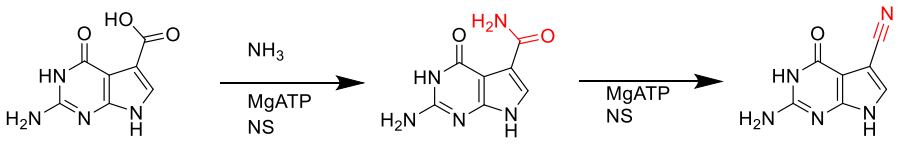


Figure 1.6: Nitrile synthetase (NS) reaction scheme. CDG is first transformed into ADG, using one equivalent of MgATP and then into $PreQ_0$ using a second equivalent of MgATP.

As the mechanism of these enzymes is not known, another aim of this work will be to gain insights on the residues involved in the transformations and engineer an amide synthetase from the NS: an engineered biocatalyst fit to carry the transformation of the carboxylic acid into an amide but not into the nitrile.

[1] B. Katz, 2018, *Smithsonian Magazine,* accessed from https://www.smithsonianmag.com/smart-news/traces-13000-year-old-beer-found-israel-180970282/ on 11/04/2020

[2] H. Fernandez, S. Hughes *et al.,* 2006, *Proc.Natl.Acad.Sci.USA,* 103, pp. 15375-15379

[3] J.P. Adams, M.J.B. Brown *et al.,* 2019, *Adv.Synth.Catal.,* 361, pp 2421-2432

[4] P.N. Devine, R.M. Howard, *et al.,* 2018, *Nature Rev.Chem.,* 2, pp 409-421

[5] A. Fryszkowska and P.N. Devine, 2020, *Curr.Op.Chem.Biol.,* 55, pp 151-160

[6] X. Huang, M. Cao and H. Zhao, 2020, *Curr.Op.Chem.Biol.,* 55, pp 161-170

[7] F.H. Arnold, 2019, *Angew.Chem.Intl.Ed.,* 58, pp 14420-14426

[8] U.T. Bornscheuer, G.W. Huisman *et al.,* 2012, *Nature,* 485, pp 185-194

[9] L. Sedlaczek, 1988, *Crit.Rev.Biotechnol.,* 1988, 7, pp 187-236

[10] H. Griengl, H. Schwab *et al.,* 2000, *Trends Biotechnol.,* 18, pp 252-256

[11] P.T. Anastas and J.C. Warner, 1998, *Green Chemistry: Theory and Practice,* Oxford University Press

[12] American Chemical Society – Green Chemistry Institute, 2020, accessed from https://www.acs.org/content/acs/en/greenchemistry/principles/12-principles-of-green-chemistry.html on 11/04/2020

[13] E.M.M. Abdelraheem, H. Bush *et al.,* 2019, *React.Chem.Eng.,* 11, pp 1878-1894

[14] J. Choi, S.S. Han and S.H. Kim, 2015, *Biotechnol.Adv.,* 33, pp 1443-1454

[15] G. Sanchez-Carron and D.J. Campopiano, 2017, *Contemporary Catalysis: Science, Technology and Applications,* Royal Society of Chemistry

[16] A. Fryszkowska and P.N. Devine, 2020, 55, pp 151-160

[17] S.K. Ma, J. Gruber *et al.,* 2010, *Green Chem.,* 12 pp 81-86

[18] Codexis, World patent application : WO2010/099501

[19] E.M. Gabor and D.B. Janssen, 2004, *Prot.Eng.Design.Select.,* 17, pp 571-579

[20] D.K. Ro, E.M. Paradise *et al.,* 2006, *Nature,* 440, pp 940-943

[21] M.A. Huffman, A. Fryszkowska *et al.,* 2019, *Science,* 336, pp 1255-1259

[22] B. Herman, 2019, *Axios*, accessed from https://www.axios.com/lipitor-pfizer-drug-patent-sales-2019-6937cdfb-47f1-46bc-8cf0-39e6b88e235e.html on 13/04/2020

[23] M. Gibney, 2016, *FiercePharma,* accessed from https://www.fiercepharma.com/special-report/16-januvia on 13/04/2020

[24] Global Newswire, 2019, accessed from https://bit.ly/2XIXhRh on 13/04/2020

[25] PR Newswire, 2019, access from https://prn.to/2wBtwXb on 13/04/2020

[26] C. Lin, R.M. McCarty and H. Liu, 2017, *Angew.Chem.Intl.Ed.,* 56, pp 3446-3489

[27] A. Dawson, P.K. Fyfe and W.N. Hunter, 2008, *J.Mol.Biol.,* 384, pp 1353-1358

[28] N.R. Williamson, H.T. Simonsen *et al.,* 2005, *Mol.Microbiol.,* 56, pp 971-974

[29] C.E. Aroyan, A. Dermenci and S.J. Miller, 2009, *Tetrahedron,* 65, pp 4069-4084

[30] H.A. Kirst, K.A. Michel *et al.,* 1991, *Tetrahedron Lett.,* 32, pp 4839-4842

[31] M. Shi, F. Wang *et al.,* 2011, *The Chemistry of the Morita-Baylis-Hillman reaction,* RSC Publishing, Cambridge, UK

[32] M. Reetz, R. Mondière *et al.,* 2007, *Tetrahedron Lett.,* 48 pp 1679-1681

[33] P.N. Joshi, L.Purushottam *et al.,* 2016, *RSC Adv.,* 6, pp 208-212

[34] S. Bjelic, L.G Nivon *et al.,* 2013, *ACS Chem.Bio.,* 8, pp 749-757

[35] L. Jiang, H.W. Yu, 2013, *Biotechnol.Lett,* 36 pp 99-103

[36] M. Winkler, K. Dokulil *et al.,* 2015, *Chem.Biol.Chem.,* 16, pp 2373-2378

Chapter 2 Towards a biocatalytic approach for the Morita-Baylis-Hillman Reaction

2.1 Introduction

2.1.1 The Morita-Baylis-Hillman (MBH) reaction

The Morita-Baylis-Hillman (MBH) reaction is a widely used C-C bond-forming reaction in organic synthesis. It involves the base-catalysed coupling of an aldehyde and an electron-withdrawing (EW) alkene. This reaction was first developed by Ken-ichi Morita and colleagues in 1968,[1] when tricyclohexylphosphine was originally used as the catalyst. Nevertheless, this catalytic system displayed limited efficiency as the reaction afforded low yields (<25%) and had to be carried out under harsh conditions. However, in 1973, Melville Hillman and Anthony Baylis significantly improved the versatility of the Morita reaction[2] (Figure 2.1). They found that using nucleophilic nitrogen-containing bases such as 1,4-diazabicyclo[2.2.2]octane (DABCO) dramatically improved the yield (>80%). The scope of the reaction was expanded to any electron-withdrawing alkenes and high temperatures were no longer required. DABCO is still widely used today as a MBH reaction organocatalyst, although it is not stereoselective

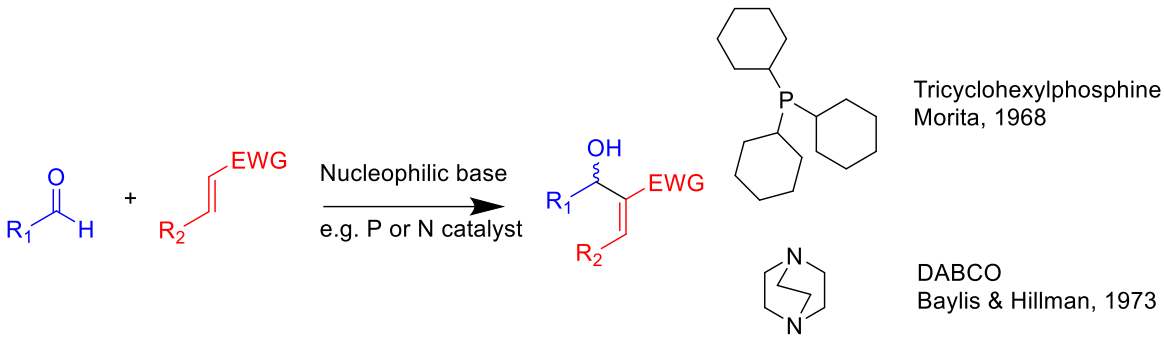


Figure: 2.1: General scheme of the MBH reaction along with the structure of the catalysts used by Morita and by Baylis and Hillman. This reaction couples an aldehyde with an electrophilic alkene to form an alcohol.

The proposed mechanism of the MBH reaction, shown in Figure 2.2, is thought to consist of three steps[3]. The first step is the Michael addition of the nucleophilic catalyst on the alkene double bond (I) to form an extremely reactive zwitterionic adduct (II). This adduct will attack the electrophilic aldehyde, forming a new carbon-carbon bond. Then a product-catalysed proton transfer, thought to be the rate-limiting step, occurs and triggers the elimination of the catalyst (III) to yield the product (IV).

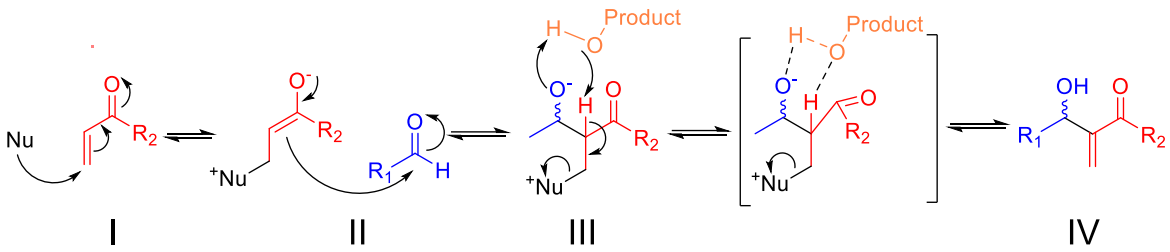


Figure 2.2: MBH reaction proposed mechanism for a vinyl ketone, an example of an electron-withdrawing alkene. The nucleophilic catalyst undergoes Micheal addition on the alkene, forming an extremely nucleophilic zwitterion which will attack the electrophilic aldehyde. A product-catalysed proton transfer and elimination of the catalyst afford the product.

As this reaction creates a chiral centre, a lot of research effort has been dedicated to control the stereochemistry of the product. The early work done in this area was concentrated on the use of unmodified chiral nitrogen bases such as brucine, quinidine or nicotine but none produced any significant stereocontrol of the reaction (<20% enantiomeric excess (ee))[4-5]. Chiral auxiliaries such as Oppolzer's sultams-(α-β) unsaturated ketone derivatives were also used, generating the corresponding 1,3-dioxan-4-one which is subsequently converted to the MBH product by transesterification with good overall yields (55-80%) and excellent ee (>99%) but vastly reduced the substrate scope, as showed in Figure 2.3[6].

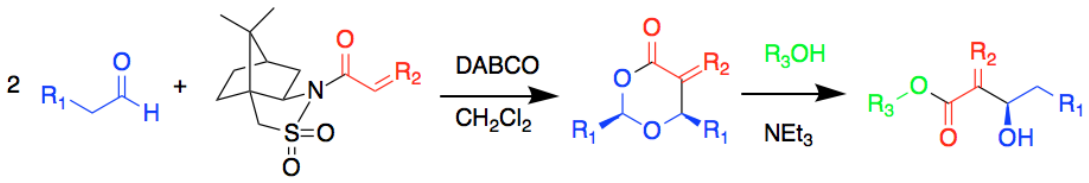


Figure 2.3: Use of Oppolzer's sultams-(α-β) unsaturated ketone as chiral auxiliaries for enantioselective MBH reaction.

The first work which reported moderate enantioselectivity (45%) for a chiral base-catalysed MBH reaction was carried by Hayase and colleagues, who used the phosphine catalyst (S)-(−)-(1,1-Binaphthalene-2,2-diyl)bis(diphenylphosphine) ((S)-BINAP) shown in Figure 2.4, but with poor yields (8-26%).

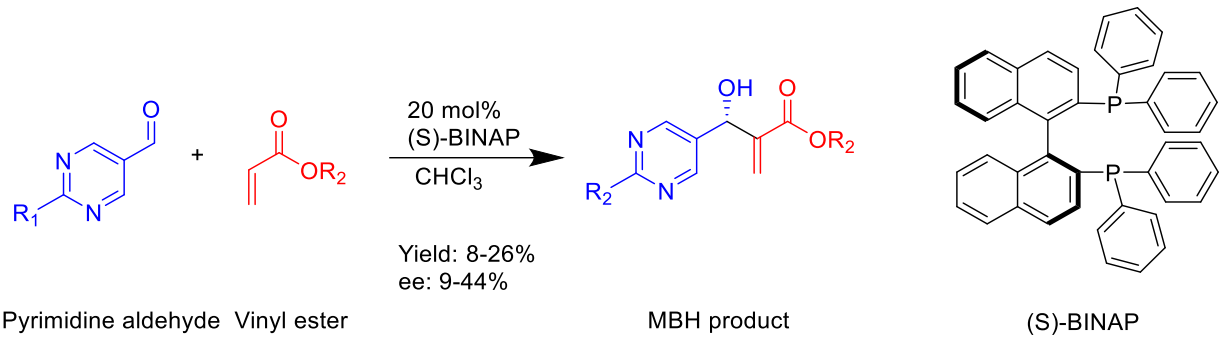


Figure 2.4: (S)-BINAP-catalysed reaction scheme and structure of the catalyst.

In 1999, however, a significant breakthrough occurred as Iwabuchi, Hatakeyama and co-workers reported a novel catalyst, thereafter called Hatakeyama catalyst, which resulted in medium to good yields (55-65%) and excellent ee (>90%) as shown in Figure 2.5[8]. The innovation in this chiral catalyst lies in the fact that, besides the Lewis basic site common to all MBH catalysts, this catalyst also features a Bronsted acidic site, which shall stabilise the enolate, direct its attack on the aldehyde and catalyse the rate-limiting proton transfer, favouring one enantiomer over the other. The weakness of this catalytic system is the formation of a few by-products (11-25%)

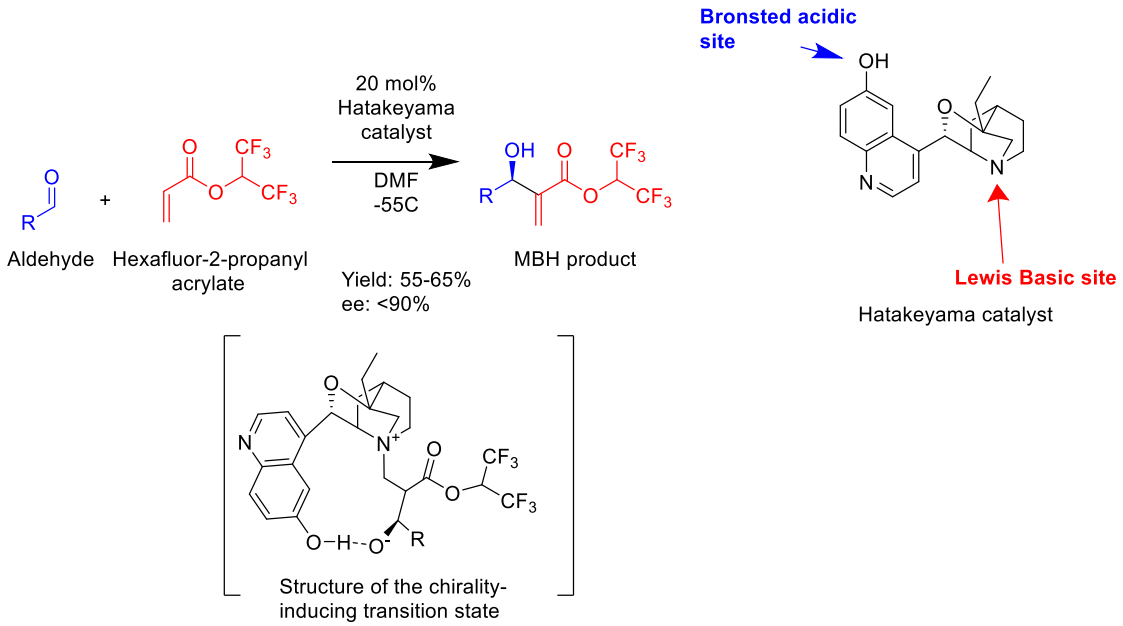


Figure 2.5: Hatakeyama-catalysed MBH reaction scheme and structure of Hatakeyama catalyst[8]. The structure of the chirality-inducing transition state is shown: one enantiomer is favoured over the other as the proton from the Bronsted acidic site of the catalyst undergoes H-bonding to the anionic alkoxyde.

The concept of complementing a Lewis basic site with a Bronsted acidic site or hydrogen bond donor to control the stereochemistry gained significant interest over the years[3]. As such, the preferred approach for organocatalyzed MBH reaction was to couple a simple, non-chiral Lewis base catalyst

with a chiral co-catalyst, which stabilise the enolate. Examples include the use of 4-dimethyl aminopyridine (DMAP) and thiourea derivatives[9] (yield: >50%, ee: >70%), or triethyl phosphine with (R)-binaphthol-derived Bronsted acids (yield: >40%, ee: >70%)[10]. This strategy was also adopted with amino acids, using proline as a Lewis base and a synthetic 7-mer peptide as co-catalyst with moderate to good yield and ee (yield: <40%, ee: 40-80%)[11].

Most recent research still relies vastly on the chiral chemical catalysts presented before and many reports have expanded their substrate scope. For example, the Hatakeyama catalyst was successfully used for the enantioselective reaction of maleimidines with 7-azaisatine derivatives to afford chiral 3-hydroxy-7-aza-2-oxindoles (yield: >40%, ee: >60%)[12] and to convert furfural and a hexafluoroisopropyl acrylate analogues for the total synthesis of the naturally-occurring antibiotic (-)-tirandamycin B (yield: 70%, ee: 99%)[13]. Binaphtalene derivatives optimised for specific substrates, resembling those presented in Figures 5 and using the same catalytic machinery have also been described recently[14-15].

Hence, as such, this reaction has many strengths:

- Unlike most carbon-carbon bond-forming reactions, which require organometallic catalysts, the use of an organocatalytic system makes this reaction cheap and environmentally benign.
- The starting materials, being an aldehyde and an alkene are usually easily accessible.
- This reaction introduces a stereogenic centre, therefore the MBH reaction can be used in a synthetic route to generate enantiopure molecules

However, this reaction has also some drawbacks:

- Even though this reaction proceeds readily with electron-withdrawing substituents, MBH reaction on electron-donating substituent is complicated as the first step, the Micheal addition of the catalyst on the alkene is not favoured.
- Although great progress was made in making the MBH reaction enantioselective, chiral control still require non-commercially available catalysts which necessitate a multistep synthesis.
- Finally, this reaction often requires very long reaction times (up to 7 days)

This aim of this workis to explore the use of biocatalysis to overcome the limitations of the chemical approach.

2.1.2 Literature review of biocatalysed MBH reactions

The most popular approach in the development of a MBH biocatalyst is to couple a nucleophilic site, which will easily attack the alkene derivative in the first step of the reaction, and a hydrogen-bond donor, which will stabilise the alkoxide intermediates, on an enzyme template. The first trial of a biocatalysed MBH reaction was published in 2007 by Manfred Reetz and co-workers, where serum albumin proteins from various species as well as some lipase enzymes were screened for MBH activity [16]. The best conversion was observed after 48h with bovine serum albumin (conversion = 35%, ee = 9%) while the best ee (conversion 15%, ee = 19%) was observed with bovine serum albumin (fraction V). as showed in Figure 2.6.

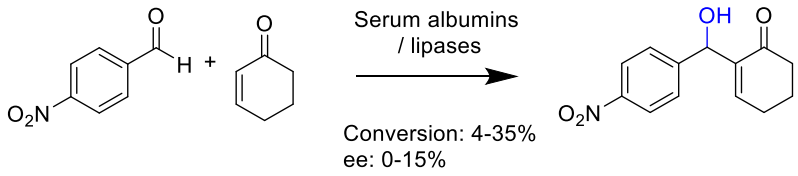


Figure 2.6: First published attempt at a biocatalysed MBH reaction by Reetz et al.[16].

Reetz & colleagues proposed a mechanism for the biocatalysed reaction: it involves a nucleophilic residue such as lysine, cysteine or histidine undergoing a Micheal addition on the alkene while the alkoxide negative charge of the subsequent intermediate is stabilised by H-bond interactions with the protein residues, in an oxyanionic hole as shown on Figure 2.7.

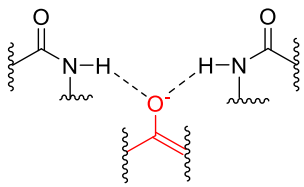


Figure 2.7: Interaction of the substrate with the protein oxyanionic hole. The substrate is stabilised by H-bond donating interactions from the protein.

In contrast, David Baker and colleagues used quantitative modelling and molecular dynamics approaches to design enone-binding biocatalysts for the MBH reaction[17]. Substantially agreeing with Reetz about the importance of H-bond donors to stabilise the oxyanionic transition states, they postulated that an efficient biocatalyst should also feature a hydrophobic and/or aromatic binding site to add subsequent favourable protein-substrate interactions. They screened 48 proteins and found two with increased activity for MBH reaction: an alanine racemase from *Geobacillus stearothermophilus* and an haloacid dehalogenase from *Pyrococcus horikoshii*. Using computational

methods, various point mutants were made and they performed saturation mutagenesis at selected positions near the active site on these two proteins. A single mutant (haloacid dehalogenase N14I), which was 2.4-fold more active than the wild-type, displayed 24% conversion after 18h, at very high protein concentration (38 mol%), with no reported ee. Notably, they were also able to observe a protein-cyclohexenone covalent adduct by mass spectrometry, providing further evidence that a nucleophilic residue attacks the electrophile.

In addition, Joshi *et al*. have screened enzymes containing a histidine and arginine pair[18]. They postulated that the histidine will act as a nucleophile while the arginine will act as a hydrogen bond donor (HBD) and stabilise the oxyanionic intermediates. Although they were able to obtain promising conversion (up to 72%) with lysozyme C from hen egg-white, they were using an extremely high catalyst loading (10 mol%).

Another approach first studied by Jiang *et al.* is to use enzymes containing a α/β-hydrolase fold (ABHF) such as lipases and hydrolases[19]. Through screening AHBF-containing enzymes, Jiang showed that pimeloyl-acyl carrier protein (ACP) methyl ester esterase (BioH) from *E. coli* (EcBioH) displayed best activity for the MBH reaction between 4-nitrobenzaldehyde (4-NBA) and methyl vinyl ketone (MVK) as 46% conversion was obtained. Considering this higher efficiency, EcBioH appears to be a promising platform to exploit to gain insights on the biocatalysed MBH reaction.

The same team have also published the EcBioH-catalysed aldol and Knoevenagel reactions and obtained up to 91% conversion for the aldol reaction and 55% for the Knoevenagel reaction which suggests an interesting synthetic promiscuity for EcBioH[20].

2.1.3 Alpha-beta hydrolase fold, an interesting structural motif

The highly conserved ABHF structural motif is characterised by 6 to 9 parallel β-sheets, which are flanked by α-helices[21]. The active site is composed of a serine protease catalytic triad made of Ser, Asp and His located on the loops that separate the helices and sheets. The function of such catalytic triad is to deprotonate the Ser to drastically increase its nucleophilicity. The pKa of serine being 15.9, non-catalytic Ser will be protonated. However, interactions with the Asp and His residues will trigger Ser deprotonation. The deprotonated Asp residue of the triad (pKa = 2.90) will give electron density through a hydrogen bond to the nearby histidine. The electron-rich histidine will, in turn, be able to activate the serine, making it capable of nucleophilic attack, as shown in Figure 2.8.

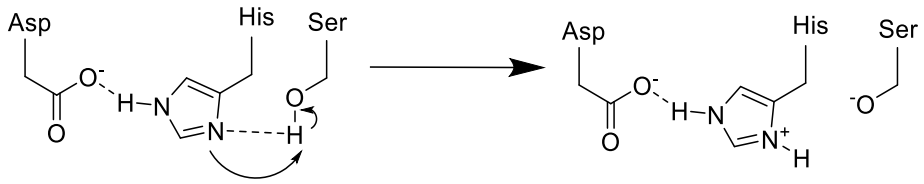


Figure 2.8: ABHF-containing enzymes catalytic triad: The nucleophilicity of the histidine is increased by the electron density given out by the deprotonated aspartic acid. This allows this histidine to abstract a proton from serine, dramatically increasing its nucleophilicity.

As such, a catalytic Ser could act as the nucleophilic activator in the MBH reaction. Furthermore, the ABHF structural motif also features an oxyanionic hole, which could stabilise anionic intermediates as shown previously on Figure 2.7, making the ABHF a promising structural scaffold for MBH catalysis. The secondary layout of the ABHF motif is shown in Figure 2.9.

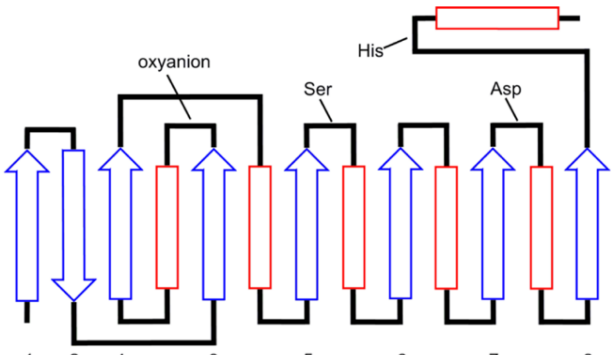


Figure 2.9: Secondary structure of the ABHF motif adapted from [16].

The proposed mechanism of the ABHF-biocatalysed MBH reaction is that the Ser in the catalytic site would attack the electrophilic alkene (I) while a hydrogen bond donor (HBD) located in the oxyanionic hole will stabilise the oxyanionic intermediate. The first intermediate, being nucleophilic will attack the aldehyde (II) and this oxyanionic intermediate will also be stabilised by the oxyanionic hole which will promote the proton transfer which will trigger the elimination of the catalyst (III), forming the product (IV) as shown on Figure 2.10.

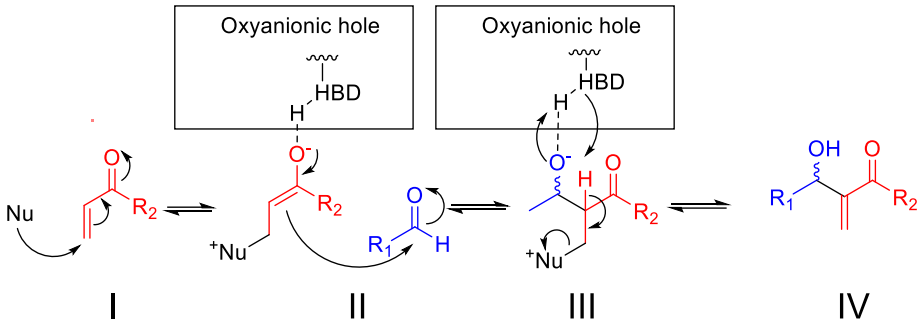


Figure 2.10: Proposed mechanism of the ABHF-catalysed MBH reaction.

2.1.4 EcBioH, a promiscuous esterase

Pimeloyl-ACP methyl ester esterase is, as its name suggests, an esterase that catalyses the hydrolysis of pimeloyl-ACP methyl ester into pimeloyl-ACP in the biotin biosynthetic pathway. In *E. coli*, it is a 29.5 kDa protein that can be found both as a monomer and as a dimer. EcBioH was first purified and characterised in the Campopiano group[22]. In *E. coli*, from malonyl-ACP methyl ester undergoes two cycles of fatty acid synthase to elongate its alkyl chain, forming pimeloyl-ACP methyl ester[26]. Then follows the BioH-catalysed hydrolysis of the pimeloyl-ACP methyl ester into pimeloyl-ACP. In effect, the BioC/BioH couple "hijacks" the fatty acid pathway. However, although most steps in this biosynthetic pathway are conserved across species, this step is catalysed by multiple enzymes, depending on the species. For most Gram-negative bacteria such as *E. coli* or *M.flagellatum,* this reaction is catalysed by BioH, which will be further discussed later[27]. However, it was shown that BioG could catalyse this reaction in *Haemophilus influenzae,* BioK in cyanobacteria, BioV in *Helicobacter,* BioJ in *Francisella* or BtsA in *Moraxella* bacteria[28]. Although all these proteins catalyse the same reaction, they share very low sequence identity (Figure 2.14), have different transcriptional regulations and belong to different subclades of the ABHF superfamily[29]. BioF then converts the pimeloyl-ACP into 8-amino-7-oxononanoic acid, with the substitution of the ACP by ethylamine[33]. Finally, BioA, BioD and BioC respectively complete the biosynthesis of biotin, as shown in Fig. 2.11.

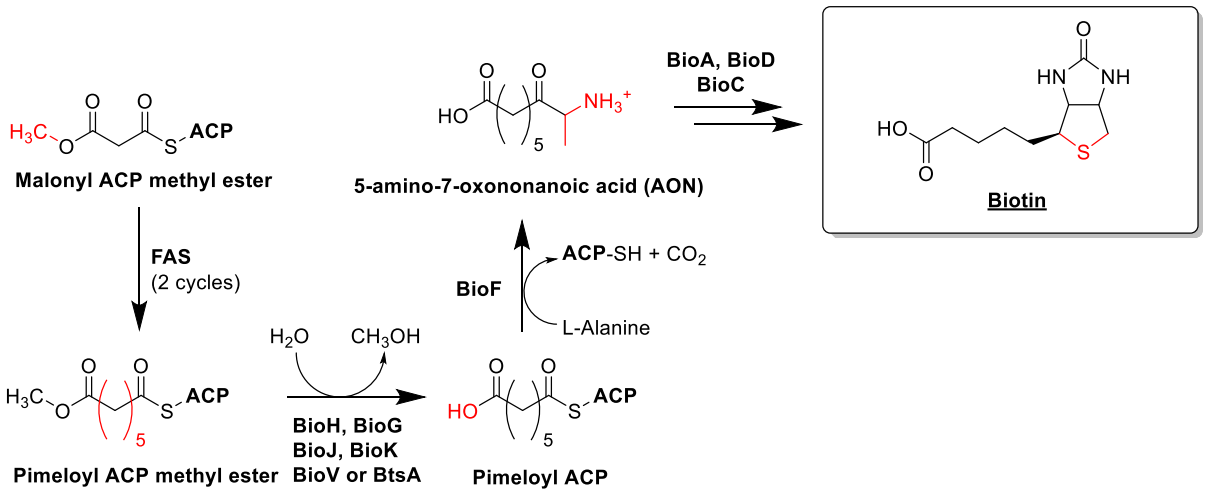


Figure 2.11: Biotin biosynthetic pathway from malonyl ACP methyl ester.
ACP: Acyl carrier protein; FAS: Fatty acid synthase

Hence, BioH role in the biotin biosynthetic pathway has been discussed. As this enzyme catalyses ester hydrolysis, it is a esterase enzyme. As discussed earlier, BioH carries a catalytic triad with a nucleophilic serine. This serine will be able to attack the substrate to catalyse the hydrolysis reaction. The methanol by-product and the substrate are then liberated, hence regenerating the biocatalyst. This mechanism is shown in Fig 2.12.

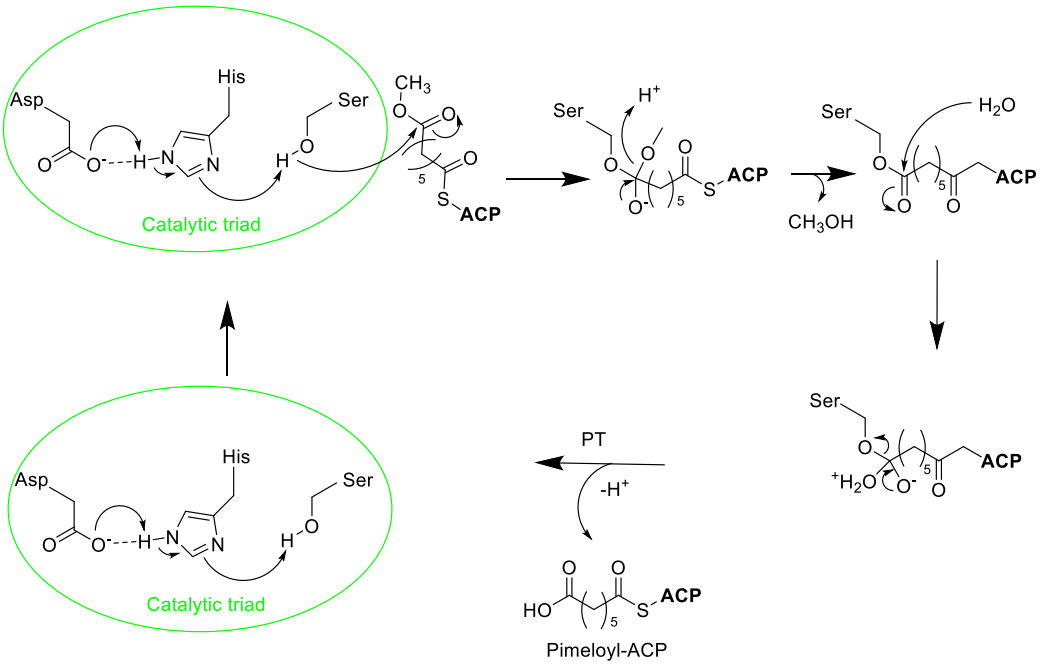


Figure 2.12: Mechanism of the BioH-catalysed hydrolysis of pimeloyl-ACP methyl ester: The catalytic serine undergo nucleophilic substitution, liberating methanol and activating the pimeloyl-ACP for the hydrolysis reaction.

BioH sequence is mostly conserved across species as displayed in Figure 2.13. However, as shown on Figure 2.14, sequence homology of EcBioH with other pimeloyl-ACP methyl ester esterases, BioG, BioV, BioJ, BioK and BtsA is very low. It was found that serine residues that are part of a conserved GX**S**XG sequence motif are likely to be involved in an esterase catalytic triad[32]. As expected, Ser82, the catalytic serine is part of a GXSXG sequence motif that is very well conserved across species. Other residues of the catalytic triad are also well conserved across species. However, interestingly, EcBioH possess a second GXSXG motif in residues 51 to 55. While this second GXSXG sequence motif is well conserved across all BioH species, it is not observed any other pimeloyl-ACP methyl ester esterases.

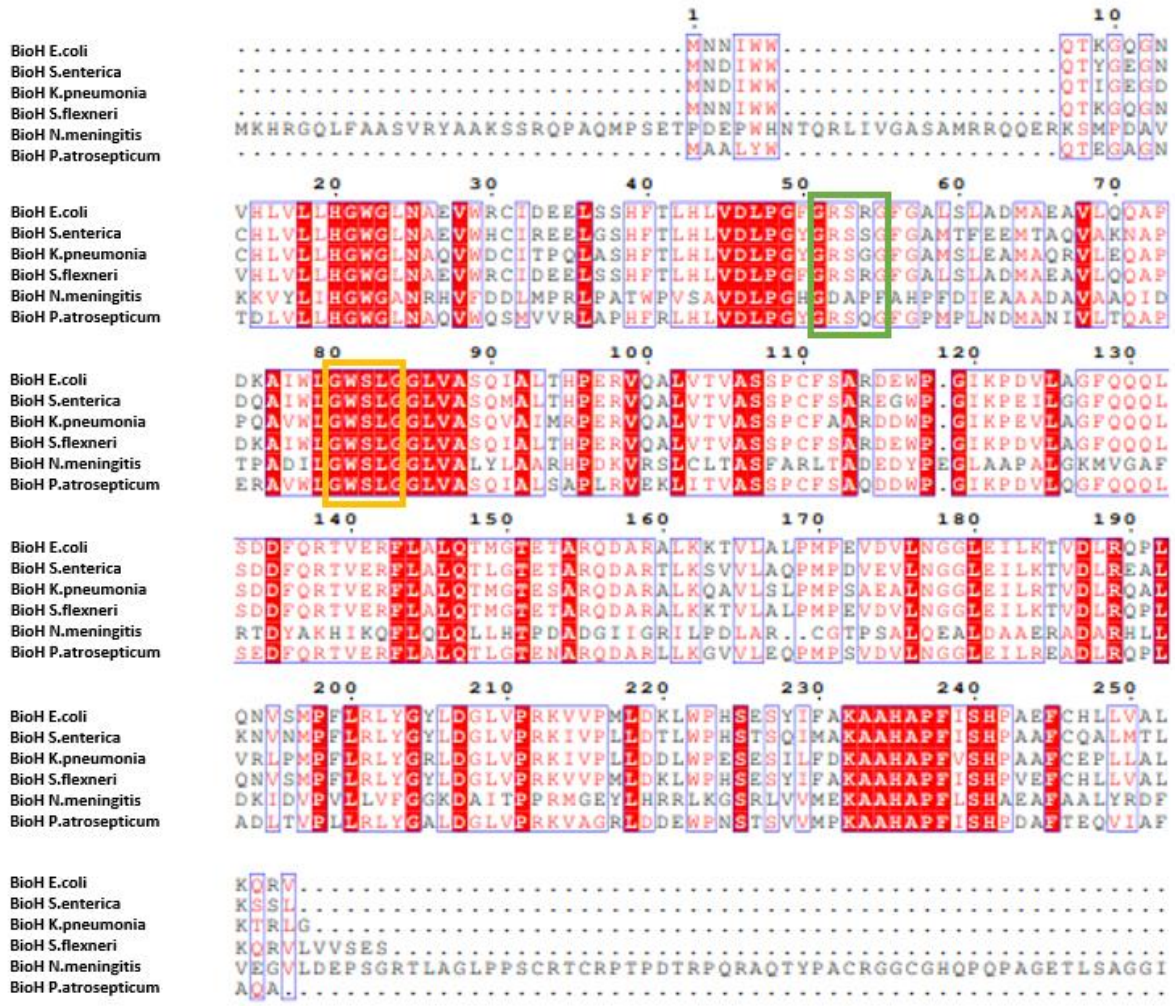


Figure 2.13: Sequence alignment between ECBioH (UniProt : P13001), Salmonella enterica BioH (A0A1R2SY91), Klebsiella pneumonia BioH (A6TF35), Shigella flexneri BioH (Q83PWO), Neisseria meningitis BioH (A0A1B1ZBF7) and P.atrosepticum (Q6CZL9) obtained from CLUSTAL omega. White over red residues are conserved across all species, red on white are conserved across most while black on white are variable across species. The yellow rectangle shows the GXSXG sequence motif thought to be part of the main catalytic site while the green rectangle shows the GXSXG sequence motif thought to be part of the cryptic catalytic site. As shown, residues involved in the GXSXG motifs and triads are well conserved across species.

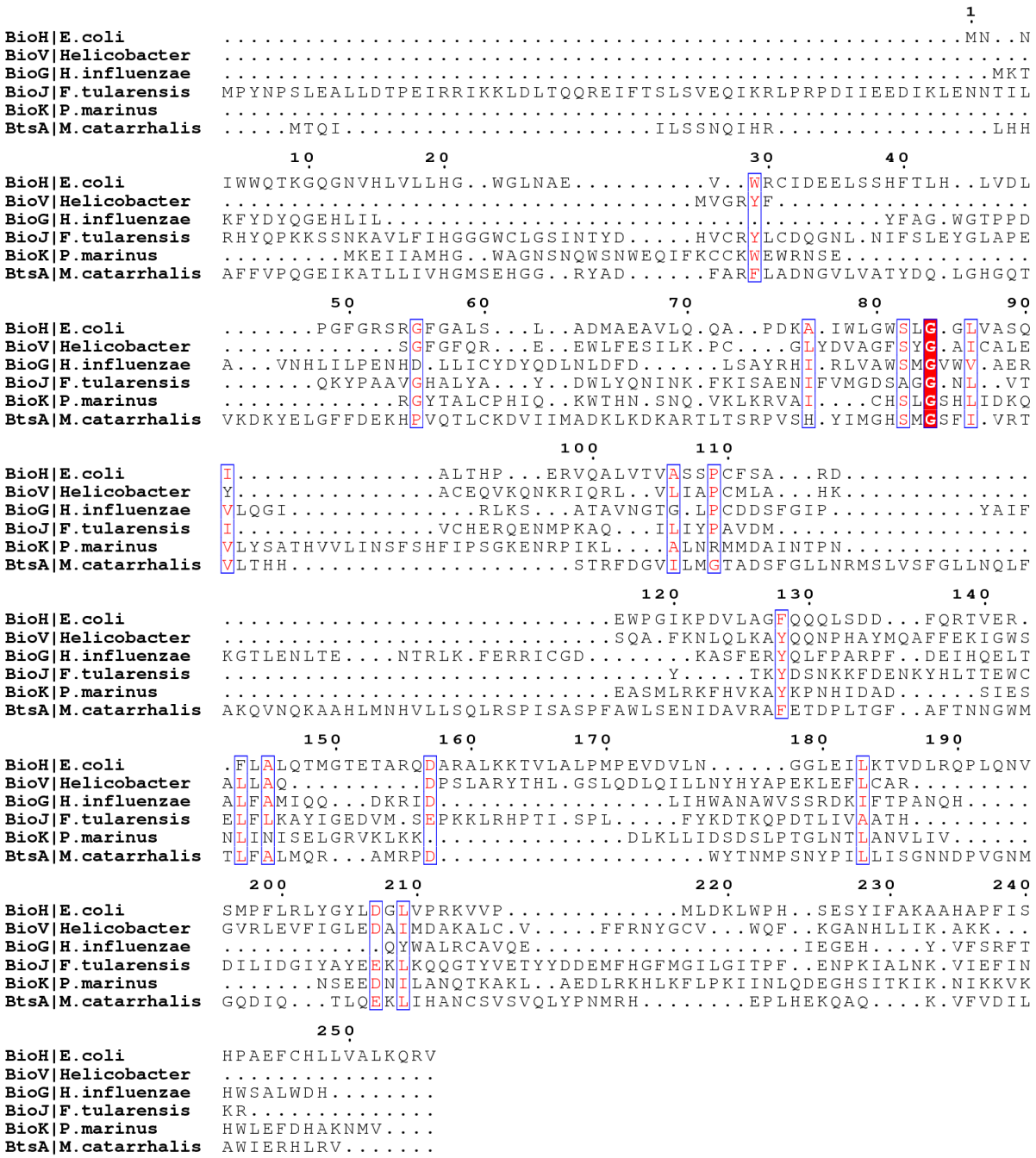


Figure 2.14 : Sequence alignment between ECBioH (UniProt : P13001), Helicobacter BioV (A0A553V3R2), H.influenzae BioG (P44251), Francisella tularensis BioJ (A0Q641), Prochlorococcus marinus BioK (A0A0A2C7Y6) and Moraxella catarrhalis BtsA (A0A3A9P6D5) obtained from CLUSTAL omega. White over red residues are conserved across all species, red on white are conserved across most while black on white are variable across species.

The structure of EcBioH was successfully determined in 2003 and is shown on Figure 2.15[30]. EcBioH was shown to have a structure typical of ABHF-containing enzymes. It presents 6 parallel β-sheets flanked by α-helices on either sides. It also displays a capping domain made of four α-helices (K121 to T185). These motifs are separated by large, flexible loops. As expected, its catalytic triad displays an arrangement which makes it extremely fit for catalysis, with His235 being within H-bond distance of both Ser82 and Asp 207. As discussed, BioH enzymes feature a second GXSXG sequence motif, suggesting the presence of a second catalytic triad, also shown on Figure 2.16.

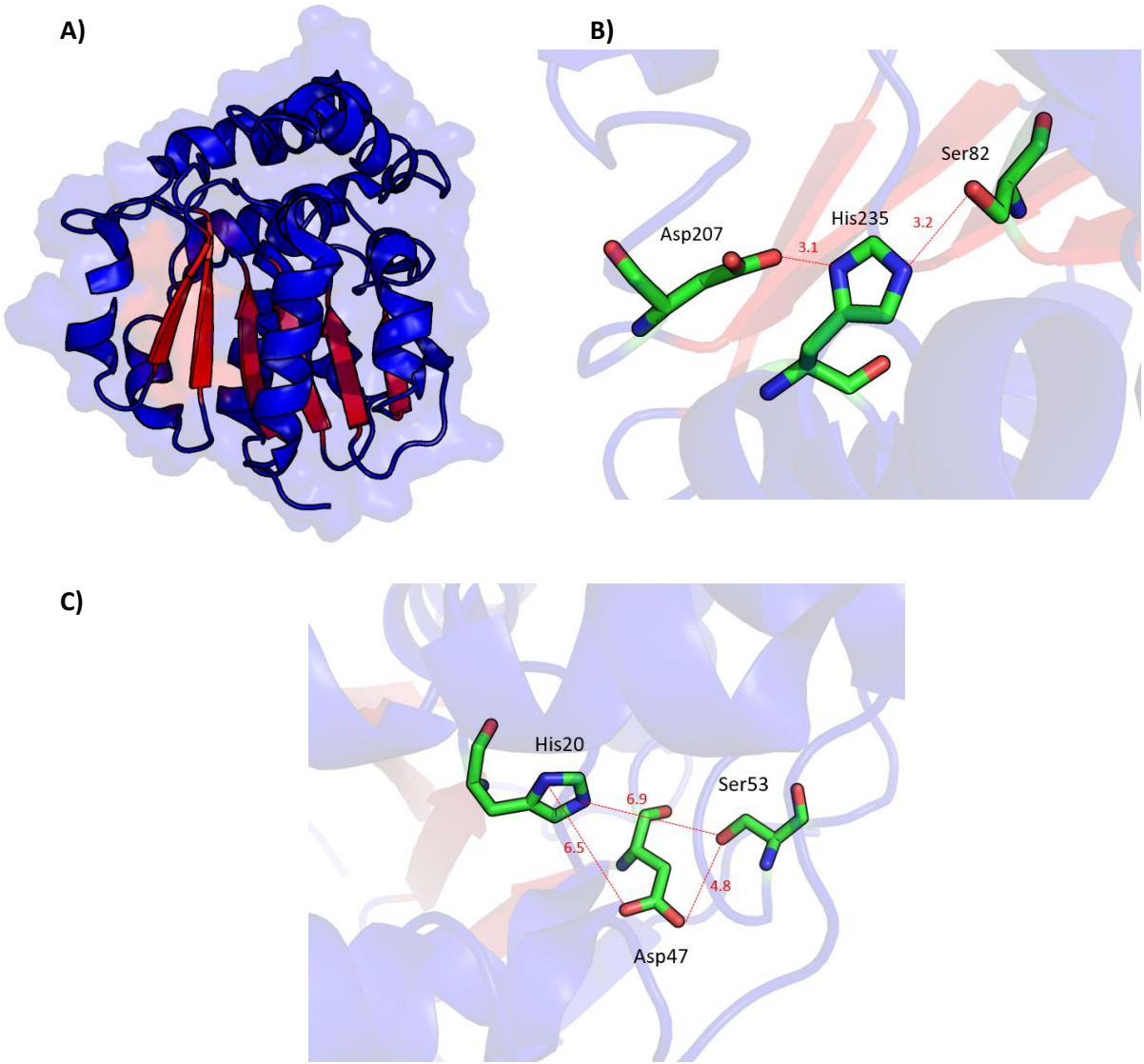


Figure 2.15 A) EcBioH WT crystal structure (PDB ID: 1M33). The ABHF motif is shown in red. B) EcBioH crystal structure zoomed on the main catalytic triad. The distance between Ser82 and His235 is 2.1Å and the distance between His235 and Asp207 is 2.2Å. C) Structure of the EcBioH potential cryptic catalytic site (PDB ID: 1M33). The distance between Ser53 and His20 is 6.9Å, and the distance between His20 and Asp46 is 6.5Å.

This cryptic site appears much less fit to catalyse the hydrolysis reaction as the intermolecular distances between Ser53/His20/Asp46 are far greater than for the Ser82/Asp207/His235 triad, reducing their ability to undergo hydrogen-bonding interaction. However, as these residues are located on loops, significant flexibility shall be expected, and they may interact together to deprotonate Ser53 and, as a result, be involved in catalysis. Other structures of BioH, from different organisms, have also been determined and are available on the PDB. It was found that they all display similar properties. These structures are summarised in Table 2.1:

PDB ID	Organism	Phenotype	Ligand	Reference
1M33	*E. coli*	Q100R, F136W (PCR errors)	None	30
4ETW	*E. coli*	S82A, V243A	Pimeloyl-ACP methyl ester	31
4NMW	*Salmonella enterica*	WT	None	Deposited on the PBD but no reference publication
6K5E	*Klebsiella pneumonia*	WT	None	33

Table 2.1: BioH crystal structures deposited to the PDB.

Interestingly, Agarwal and co-workers were able to co-crystallise EcBioH with its ligand pimeloyl-ACP methyl ester[31]. However, they were not able to co-crystallise the wild-type enzyme but instead the catalytic serine knockout mutant (S82A). While this mutation might induce a different binding mode for pimeloyl-ACP, it nonetheless gives interesting insights on the structure of the enzyme-ligand complex. As expected, the distal part of the pimeloyl moiety binds the catalytic triad, for reaction (which is not occurring in this case due to the S82A mutation). The ACP displays four helices with a hydrophobic core which undergoes protein-protein interactions with EcBioH, as shown on Figure 2.16.

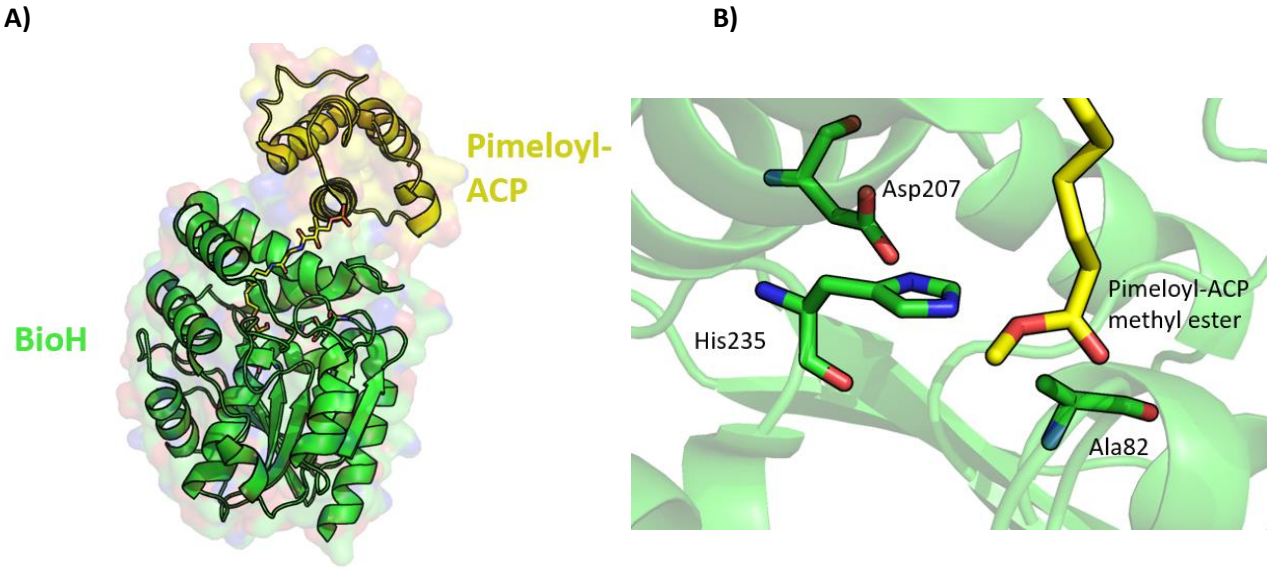


Figure 2.16: EcBioH S82A V243A co-crystal structure with a pimeloyl-ACP methyl ester (PDB: 4ETW).
A) Full structure. EcBioH is shown in green. Pimeloyl-ACP is shown in yellow.
B) Zoomed on the catalytic triad. EcBioH is showed in green, pimeloyl-ACP methyl ester is showed in yellow. The residues of EcBioH catalytic triad and the pimeloyl methyl ester moiety and the ACP phosphopantetheine arm are shown as sticks.

In addition to BioH, the crystallographic structures other pimeloyl-ACP methyl ester esterases were determined. BioG from *H.influenzae* structure was determined by the group of John Cronan in 2016[34] (PDB ID: 5GNG) and BioJ structure was determined in 2019, although the structure has still not been made available on the PDB (PDB ID: 6K1T)[27]. All the published structures of BioH enzymes were overlapped as well as the structure of EcBioH with *H.influenzae* BioG, as shown on Figure 2.17.

A) B)

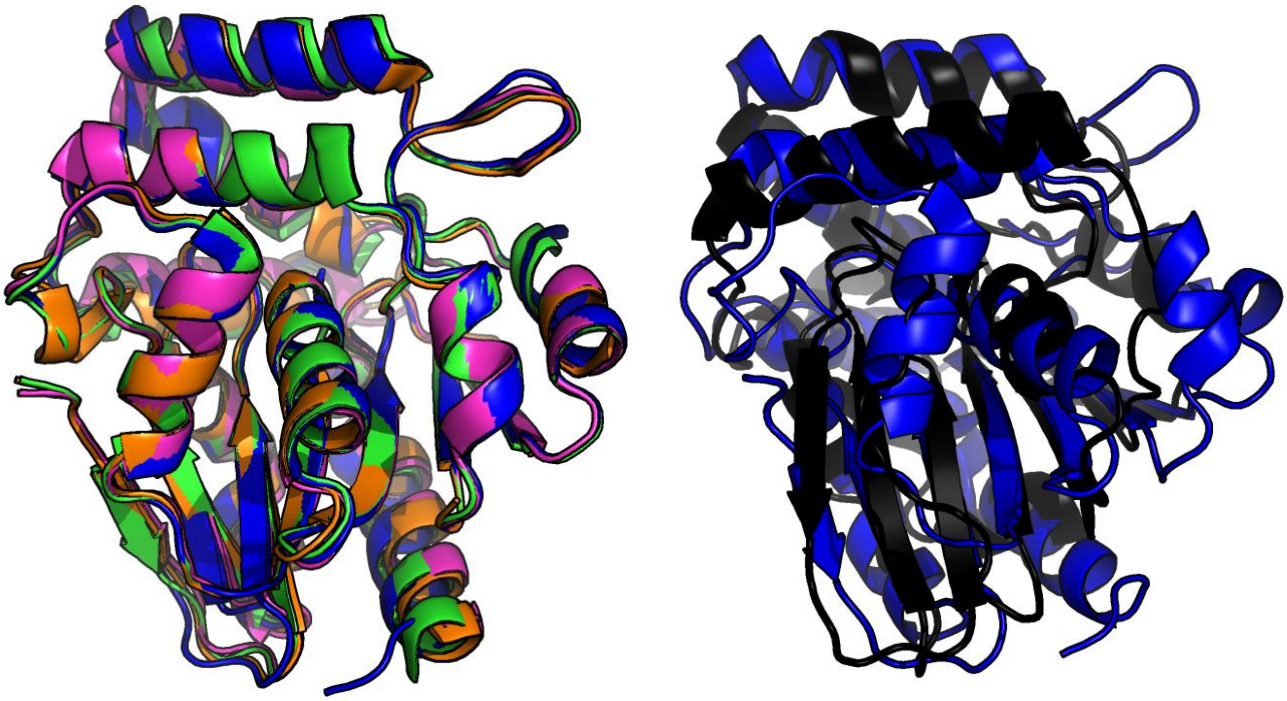


Figure 2.17: A) Overlay of BioH structures: E. coli BioH: PDB ID: 1M33 (blue) and PDB ID: 4ETW (green); Salmonella enterica PDB ID: 4NMW (magenta) and K.pneumonia PDB ID: 6K5E (orange). B) Overlap of EcBioH structure PDB ID 1M33 (blue) with H.influenzae BioG PDB ID: 5GNG (black).

From Figure 2.18, it was observed that BioH displays a very high degree of structural similarity across species. In contrast, BioG is structurally more different and displays an atypical ABHF. While on BioH, the parallel β-sheets are flanked by α-helices on both sides, in BioG, large loops replace four of six of these α-helices. The parallel β-sheets core is, however, very similar in BioH and BioG.

To obtain detailed insights into the biocatalysed MBH reaction mechanism, the role of all the features discussed will be explored. In first instance, EcBioH catalytic triad will be deactivated. This would determine whether the catalytic triad is involved in MBH catalysis. Thus, if a normal or slightly reduced activity is observed when the catalytic triad is inactivated, it will be possible to conclude that another feature of the enzyme plays a predominant role in the reaction. Such features will be investigated.

2.1.5 Aims

The general objective of this part is to gain mechanistic insight into the biocatalysed MBH reaction. As such, the aims of this project are:

- To clone, express, purify and characterise EcBioH.
- To design a suitable assay to monitor efficiently and precisely the MBH reaction.
- To optimise conditions for the biocatalytic MBH reaction.

- To deactivate the conserved and cryptic catalytic triads by site-directed mutagenesis and to assess the role of the catalytic triads in the MBH reaction.
- To gain mechanistic insights of the MBH reaction.
- To screen other proteins and enzymes, which could catalyse the MBH reaction.

2.2 Generation of the biocatalyst

2.2.1: Cloning of *E. coli* BioH

As discussed in the introduction, the starting point of this study is the esterase BioH from *E. coli* (EcBioH) as this enzyme was found by Jiang and co-workers to be able to catalyse the Morita-Baylis-Hillman (MBH) coupling of 4-NBA and MVK with 46% conversion (Figure 2.18[20]).

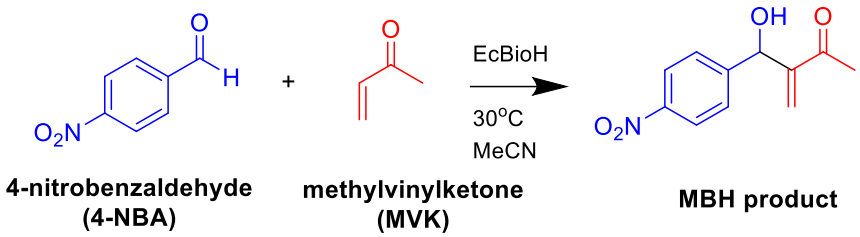


Figure 2.18: The EcBioH-catalysed MBH reaction scheme between 4-NBA and MVK.

As a first step, the full-length *BioH* gene (*E. coli* strain K12, UNIPROT code: P13001) was cloned in different expression vectors and with or without the presence of a stop codon, in order to give two different recombinant forms of EcBioH: one with a non-cleavable N-terminal hexahistidine tag and one with a non-cleavable C-terminal hexahistidine tag. More specifically, the gene was cloned in pET22b and pET28a plasmids. Both plasmids possess two interesting features which make them particularly suitable for the overexpression of proteins:

- An antibiotic resistance gene (kanamycin for pET28a and ampicillin for pET22b) which allows cells containing the plasmid to be resistant to a selected antibiotic, which will be added to the growth media to minimise the risk of contamination.
- A *lac1* operon, which allows the use of isopropyl-B-D-thiogalactopyranoside (IPTG) as an inducer for enzyme overexpression. IPTG will allosterically bind and inactivate the lac repressor, hence allowing RNA polymerase to initiate T7 transcription.

All these variants were cloned with Nde1 and Xho1 restriction sites, allowing easy transfer in different plasmids if required. The basic information of these variants of EcBioH are shown in Table 2.2:

Entry name	Position of the hexahistidine-tag	Number of amino acids	Vector	Stop codon
EcBioH-N	N-terminal	276	pET28b	Yes
EcBioH-C	C-terminal	264	pET22b	No

Table 2.2: Table showing the different recombinant forms of EcBioH cloned.
** Molecular weights were calculated by ExPASy-ProtParam tool. Average mass is showed.*

EcBioH-C from Table 2.2 was cloned by Dr Menglu Wang while EcBioH-N was cloned at the commencement of this study. The vector maps for these two entries are shown in the appendix.

2.2.2: Expression and purification of EcBioH.

All variants of EcBioH were expressed in BL21 (DE3) *E. coli* strain after induction with 0.3 mM IPTG and incubation at 18ºC for 18 h, following the conditions presented by Agarwal and Cronan, which gave satisfactory results[31]. The enzyme was purified by immobilised Nickel affinity chromatography (IMAC) followed by size-exclusion chromatography (SEC). The hexahistidine-tagged EcBioH were purified with a gradient from 30 to 250 mM imidazole. Both EcBioH-N and EcBioH-C eluted at 75 mM imidazole. SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) analysis showed that the protein was recovered with a purity of 80+ %, with traces of a few higher molecular weight proteins as shown on Figure 2.4. The SEC step was initially carried on a Superdex S200 column (120 mL). Interestingly, two large peaks were observed in every case: one with an elution volume of 76 mL and the second one with an elution volume of 84 mL. Analyses in denaturating conditions (SDS-PAGE and LC-MS-(ESI)) showed that the mass of the monomer in both cases was the same (approx. 30 kDa), suggesting that two different EcBioH quaternary structures were obtained. According to the Superdex S200 column calibration curve, the first peak corresponds to a mass close to the dimeric EcBioH (Determined Mwt of the complex from S200 calibration curve: 58 kDa; Expected Mwt of EcBioH dimer: 59 kDa) while the second peak should correspond to monomeric EcBioH (Determined Mwt of the complex from S200 calibration curve: 26 kDa; Expected Mwt of EcBioH monomer: 29 kDa). However, the separation of these peaks was not ideal. After investigating this issue, it was found that when using a Superdex S75 column two better resolved peaks were obtained with elution volumes of 54 mL and 66 mL. In addition, one peak corresponding to higher molecular weight is observed, as shown in Figure 2.19 for EcBioH-C.

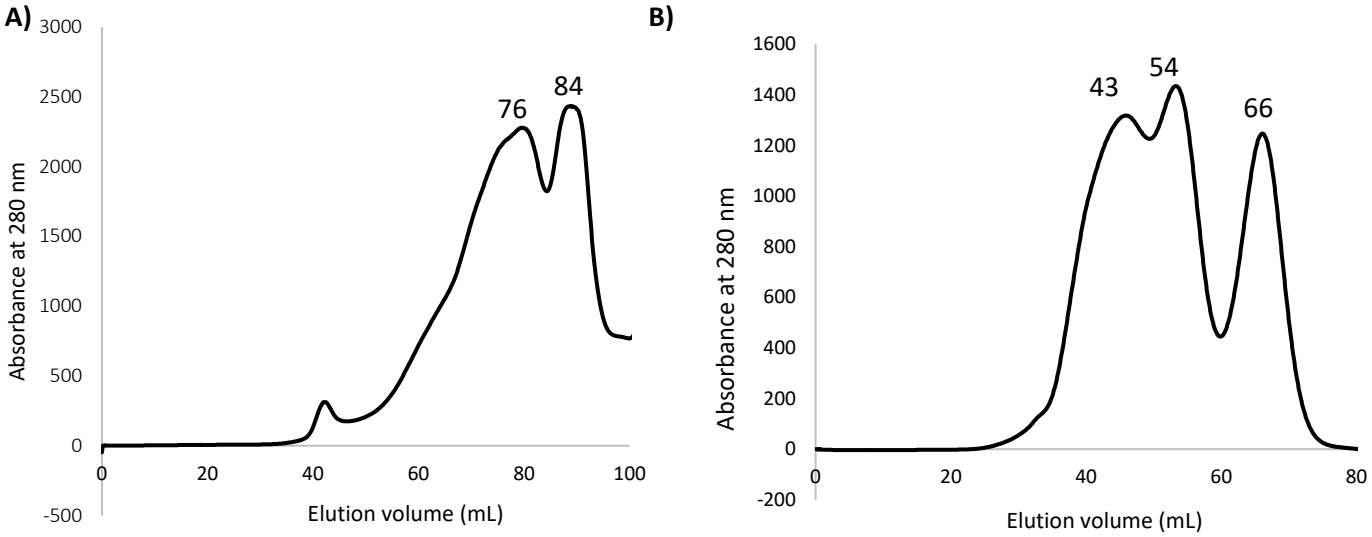


Figure 2.19: Gel filtration chromatograms of purified EcBioH-C.
A): Superdex S200 shows one small peak at 40 mL corresponding to high molecular weight aggregates, one peak at an elution volume of 76 mL corresponding to EcBioH-C dimer and one peak at an elution volume of 84 mL corresponding to EcBioH-C monomer.
B): S75 chromatogram, showing one peak at 42 mL corresponding to high molecular weight EcBioH-C aggregates, one at 54 mL corresponding to the dimer and one at 66 mL which corresponds to EcBioH-C monomer. S75 purification affords the best resolution.

It was investigated whether it is possible to control the yields of dimeric and monomeric versions of EcBioH. Indeed, while it was found that these quaternary structures display similar activities on the various assays carried, it was thought to be important to have a single quaternary structure for structural studies. While changing the buffer didn't significantly affect the distribution, it was found that lowering the flow rate from 1 mL/min to 0.6 mL/min favours the isolation of the monomer as shown on Figure 2.20 for EcBioH-C.

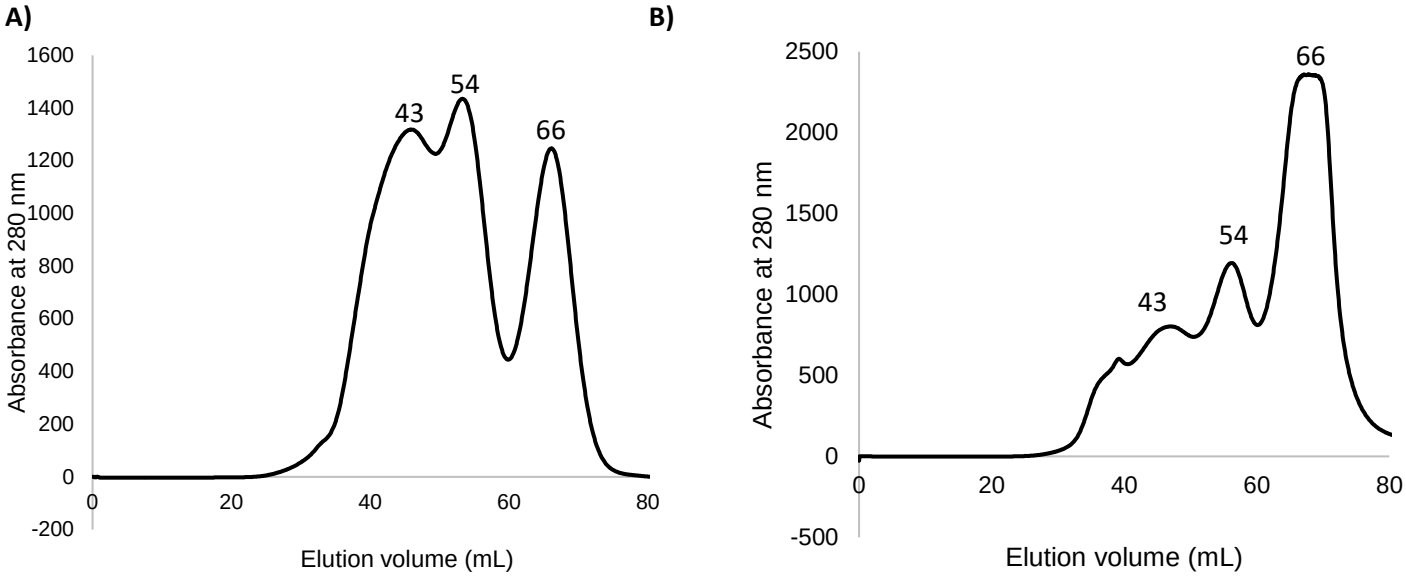


Figure 2.20: Superdex S75 chromatograms of EcBioH-C.
A): flow rate = 1 mL/min
B): flow rate = 0.6 mL/min. The peak at 65 mL was found to be much larger upon lowering the flow rate, suggesting increased purification of EcBioH-C monomer.

While baseline resolution of the monomer and the dimer was not obtained, the monomer was recovered with high purity when the flow rate was reduced, with the dimer present as a small impurity. The impact of the size-exclusion chromatography flow rate on protein behaviour was already described by Harrowing and Chaudhuri on the size-exclusion refolding of β-lactamase[35].

From SDS-PAGE analysis, it was found that after size-exclusion chromatography, the enzyme purity increased to 95+% for EcBioH-C. A SDS-PAGE gel is shown in Figure 2.21.

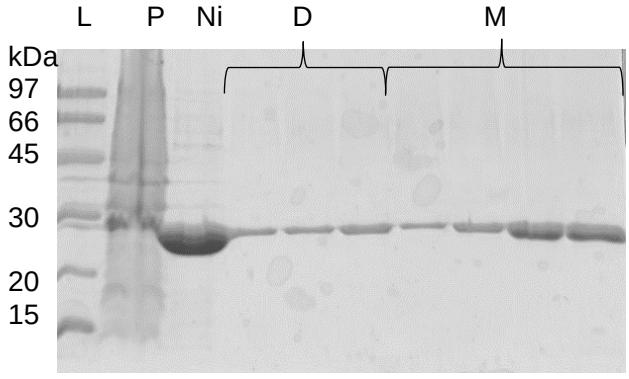


Figure 2.21: 12% SDS-PAGE gel of EcBioH-C purification steps
L: Low molecular weight ladder; P: Cell pellets; Ni: After immobilised nickel affinity chromatography; D: Fractions corresponding to the dimer after size-exclusion chromatography; M: Fractions corresponding to the monomer after size-exclusion chromatography.

Yields of 10-15 mg/L of culture of pure EcBioH were routinely obtained. The constructs were characterised by electrospray ionisation mass spectrometry (ESI-MS). A routine MS spectrum for EcBioH is showed in Figure 2.22. Low levels of impurity which could be the oxidised protein were observed. The theoretical and measured masses for the different EcBioH are shown in Table 2.3:

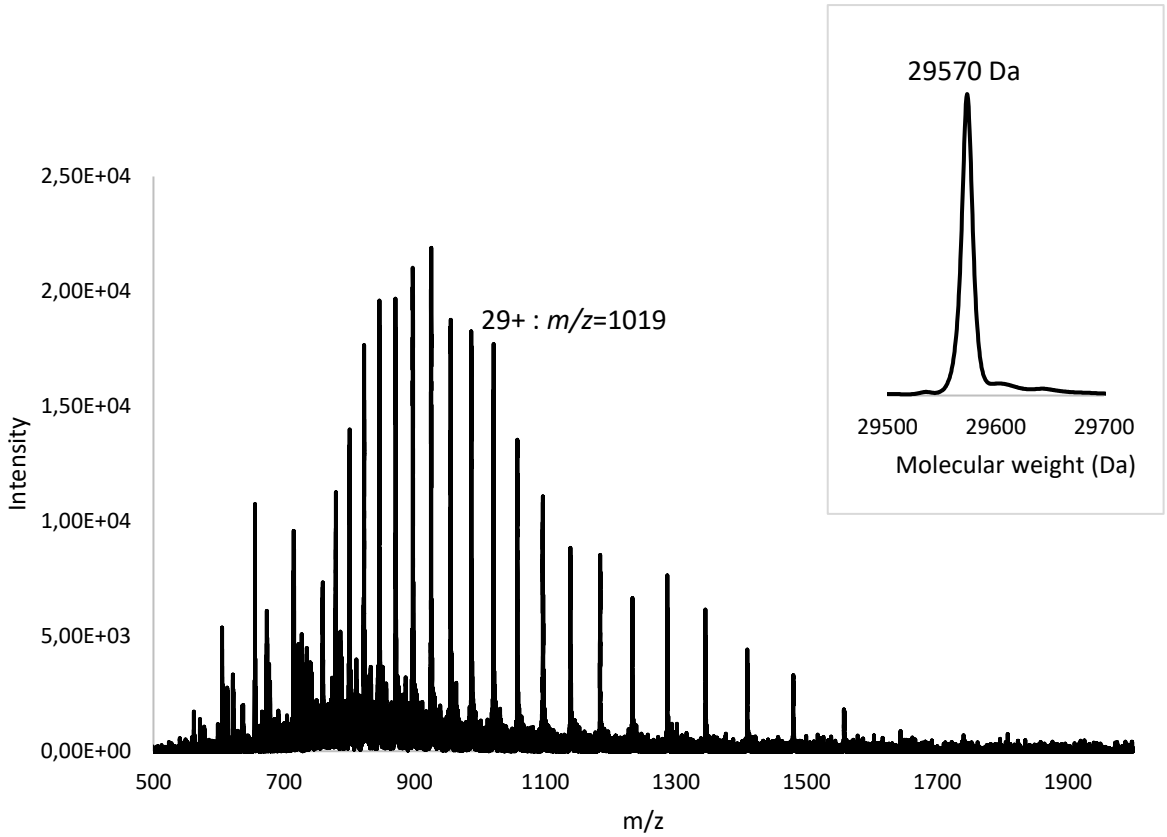


Figure 2.22: ESI-MS spectrum for EcBioH-C. EcBioH-C deconvoluted mass spectrum is shown.

Entry number	Position of the hexahistidine-tag	Theoretical molecular weight*	Experimental molecular weight
1	N-terminal	30668.34 Da	30536.51 ± 2.37 Da
2	C-terminal	29570.13 Da	29569.61 ± 1.48 Da

Table 2.3: MS data of the first EcBioH clones expressed.
** Theoretical molecular weights were calculated by ExPASy-ProtParam tool. Average mass is showed.*

2.3 Development of robust assays

Before optimising the biocatalysed MBH reaction, a method to efficiently monitor the reaction was required. The MBH product was obtained synthetically and fully characterised by nuclear magnetic resonance (NMR) spectroscopy and MS (See Methods). Hence standards for each reagent as well as the product were available.

2.3.1 A robust, medium-throughput reverse phase HPLC MBH assay

A useful tool to quantitatively assess the progress of a chemical reaction is high-performance liquid chromatography (HPLC). With this method, the components of a reaction mixture are separated on a column and subsequently detected by an ultraviolet (UV) detector. Quantitative data can easily be obtained as the area of the peak will depend on the extinction coefficient of the compound at that wavelength, and its concentration following the Beer-Lambert Law. Conditions which resulted in baseline resolution of the standard MBH reaction mixture were found to be an isocratic elution of 50% MeCN – 50% water at 1 mL.min^{-1} with UV detection at 254 nm. This wavelength was specifically chosen because it is the wavelength at which phenyl rings show maximum absorption[36]. In addition, methyl vinyl ketone does not absorb at this wavelength. Hence, what was eventually observed was the conversion of the 4-nitrobenzaldehyde into the MBH product. Consequently, methyl vinyl ketone was always used in excess. The HPLC chromatogram obtained when mixing 4-NBA, MVK and the MBH product is shown in Figure 2.23 along with the standard reaction monitored in this study. Calibration curves for 4-NBA and the MBH product are shown in the appendix.

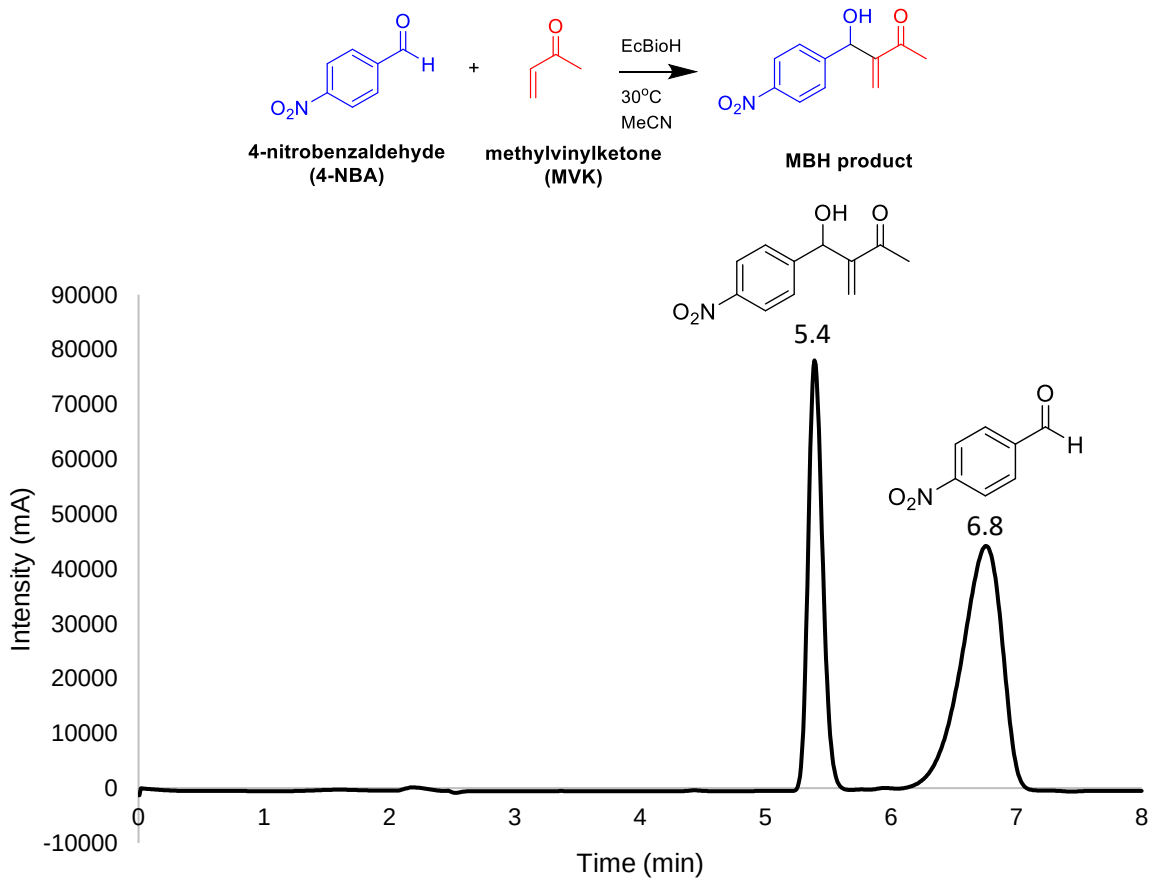


Figure 2.23: HPLC analysis of an equimolar mixture of 4-NBA, MVK and the corresponding MBH product. The peak with retention time (Rt) = 5.4 min corresponds to the MBH product while the peak with Rt = 6.8 min corresponds to 4-NBA. Methyl vinyl ketone does not absorbs at this wavelength. Calibration curves with 4-nitrobenzaldehyde and the MBH product were drawn (Appendix) and were used to determine the extinction coefficients of the two species: $\varepsilon_{nitrobenzaldehyde}$ *= 7326 L/mol/cm and* $\varepsilon_{MBH\ product}$ *= 8126 L/mol/cm.*
Conditions: Luna 5µ C18 (2) RP-HPLC column (100Å, 250×4.60mm, Phenomenex) eluting at 1 mL/min in water/MeCN 1:1 + 0.1% trifluoroacetic acid (TFA) (isocratic).

2.3.2 Chiral HPLC MBH assay

As the MBH reaction induces chirality, a way to distinguish and quantify both enantiomers of the MBH product was investigated. A chiral HPLC method was therefore set up and optimised for high resolution separation of both enantiomers. Stationary and mobile phases were therefore screened. It was found that the method which led to the best enantiomeric resolution was an isocratic elution of 8% isopropanol in hexane using a ChiralPak IC column. This column uses cellulose tris(3,5-dichlorophenylcarbamate) as a chiral selector. Both enantiomers interact differently with this stationary phase, hence allowing chiral separation. As previously, the wavelength for detection was 254 nm. The chiral HPLC chromatogram of a racemic mixture of MBH product is shown in Figure 2.24.

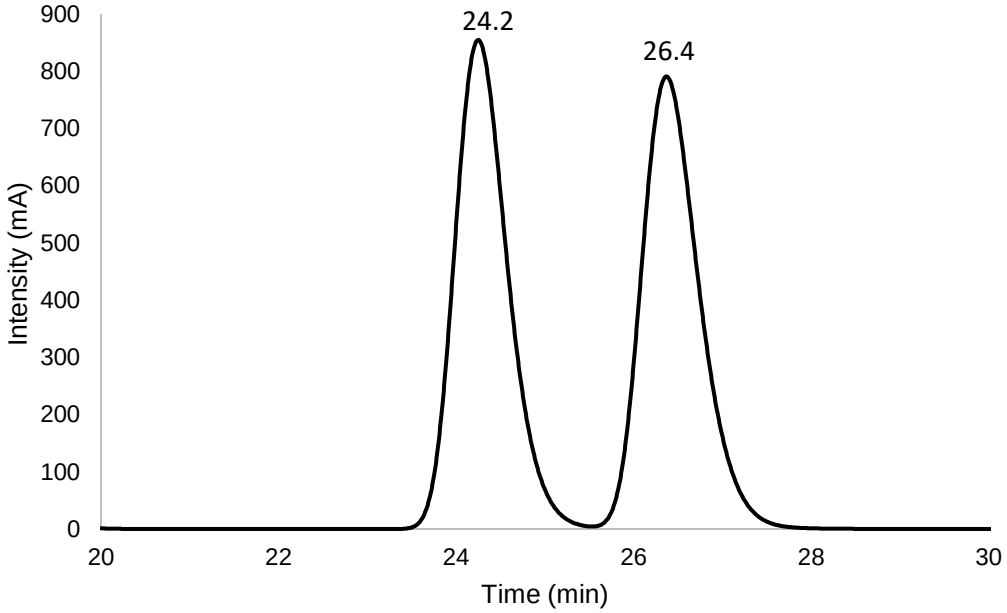


Figure 2.24: Chiral HPLC chromatogram of a racemate of the MBH product at 254 nm. The retention times of the enantiomers were found to be 24.2 min and 26.4 min. The chirality of the species corresponding to each peak is unknown.
Conditions: ChiralPak IC column eluting at 1 mL/min in 8% isopropanol in hexane

Although this method had not been optimised as previously to quantify the conversion of 4-NBA into the MBH product, it was verified that the 4-NBA doesn't co-elute with any of the enantiomers.

2.3.2 Investigating alternative high throughput MBH assays

Although an efficient HPLC methodology was set up, the required sample preparation and the lengthy analysis make it impractical for high-throughput (HTP) assay of the MBH reaction. Thus, a spectrophotometric assay had to be developed to allow the simultaneous monitoring of up to 96

reactions using a 96-wells plate reader. More precisely, an assay detecting either aldehydes or alcohols, but not ketones had to be developed. A Schiff reagent assay was therefore trialled. In this assay, it is expected that 4-NBA will react with a sulphurous fuschin derivative to form a purple product following the scheme shown in Figure 2.25.

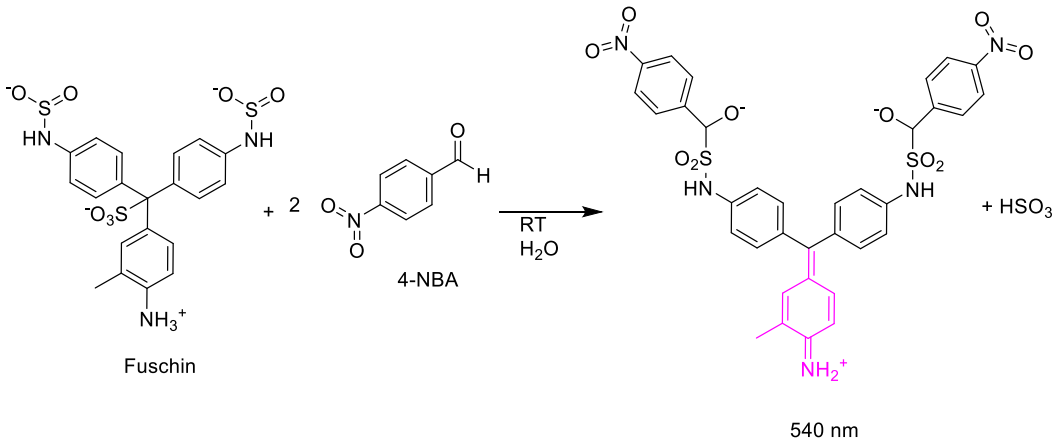


Figure 2.25: Schiff reagent assay scheme.

Hence, the Schiff reagent will selectively react with aldehydes to form an adduct which displays high UV absorbance at 540 nm, as shown in Figure 2.26.

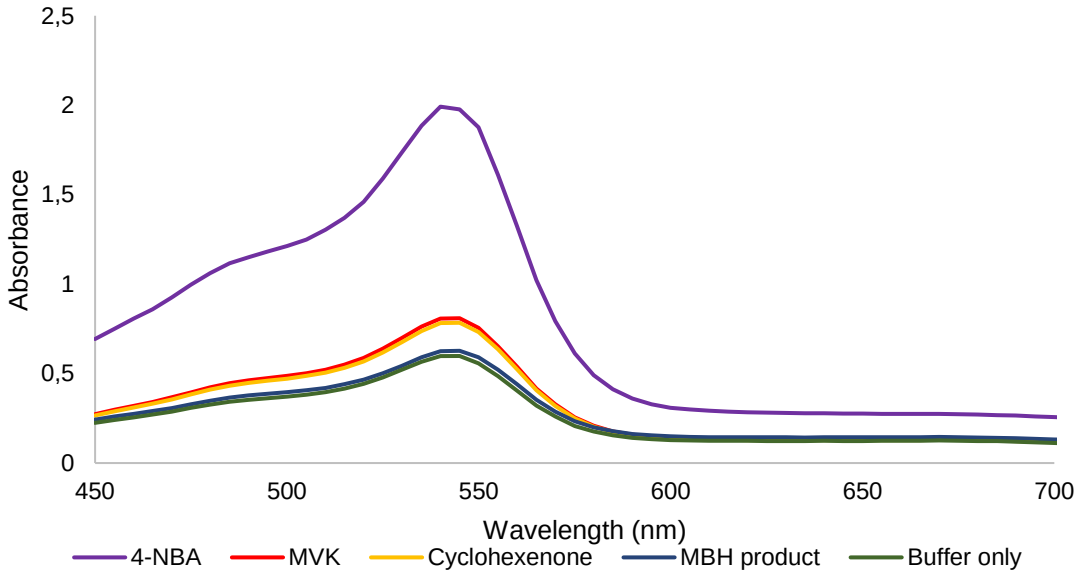


Figure 2.26: UV-Vis spectrum of Schiff reagent in mixture with 4-NBA, methyl vinyl ketone (MVK), cyclohexenone, the MBH product and the buffer between 450 and 700 nm after one minute incubation. A significant absorbance at 540 nm is observed for 4-nitrobenzaldehyde.

Analysis by UV-Vis spectrometry shows that this assay is selective towards 4-NBA. Unsurprisingly, α-β unsaturated ketones such as methyl vinyl ketone and cyclohexenone also show little conversion as small ketones can also react with this reagent. As the MBH product is bulky, no

cross-reactivity was observed with Schiff reagent. However, the incubation time had to be closely monitored (maintained below one minute) as, after 30 minutes, all the reagents, the buffer alone or deionised water, incubated with Schiff reagent showed a deep purple colour and a large peak at 540 nm on UV-Vis spectroscopy. This suggests that Schiff reagent is very reactive and that it would, therefore, be challenging to control background reactions. Besides, after multiple attempts in creating a calibration curve, no satisfactory linearity was achieved on the area under the curve (AUC) vs concentration graph. Reasons for this issue are still unclear as the mechanism of the reaction between the aldehyde and Schiff reagent is unclear.

From these issues, it was concluded that the most effective way to monitor the conversion efficiently was the HPLC assay presented in part 2.3.1.

2.3.4 Esterase reaction assay

This project aims to take advantage of the EcBioH esterase promiscuity to generate biocatalysts which would work efficiently for the MBH reaction. However, the natural esterase activity of the enzyme should also be assayed as it would be interesting to compare how the esterase activity is affected by the enzyme engineering targeting the MBH catalytic activity. A substrate which undergoes a change in UV absorbance upon hydrolysis would be ideal for this as conversion could be easily monitored on a 96-well plate reader.

p-nitrophenyl acetate (pNPA) was found to be interesting on that regard since its hydrolysis product, *p*-nitrophenol (pNP), absorbs at 410 nm when deprotonated. The scheme of that hydrolysis reaction is shown in Figure 2.27.

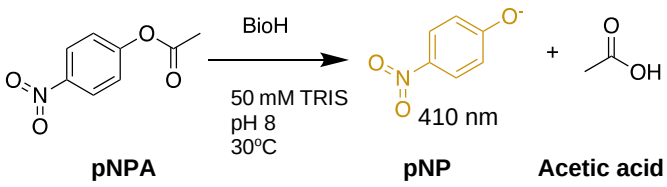


Figure: 2.27: p-nitrophenyl acetate assay reaction scheme.

Hence, various concentrations of pNPA were reacted with EcBioH and formation of *p*-nitrophenol was monitored every 15 seconds for one hour. Figure 2.28 shows this conversion in the first 290 seconds for EcBioH-C.

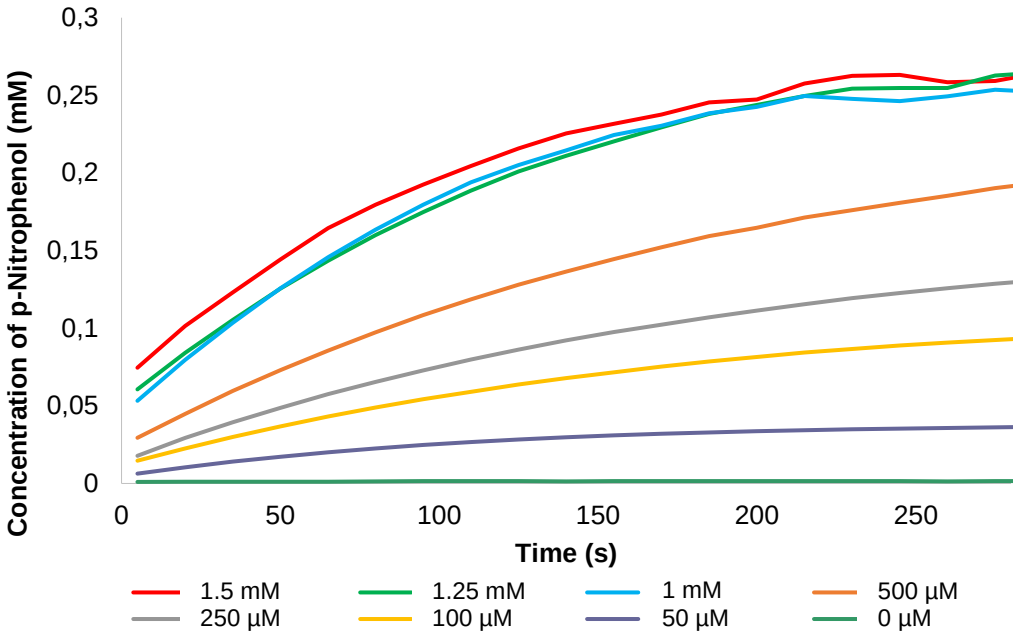


Figure 2.28: Formation of p-nitrophenol monitored over time (first 290s shown). Enzyme concentration: 67 nM

The initial velocity of the enzyme at each concentration was then found by calculating the gradient for the linear part of the curves (0 to 80 s) and the Michaelis-Menten plot was drawn (Figure. 2.29).

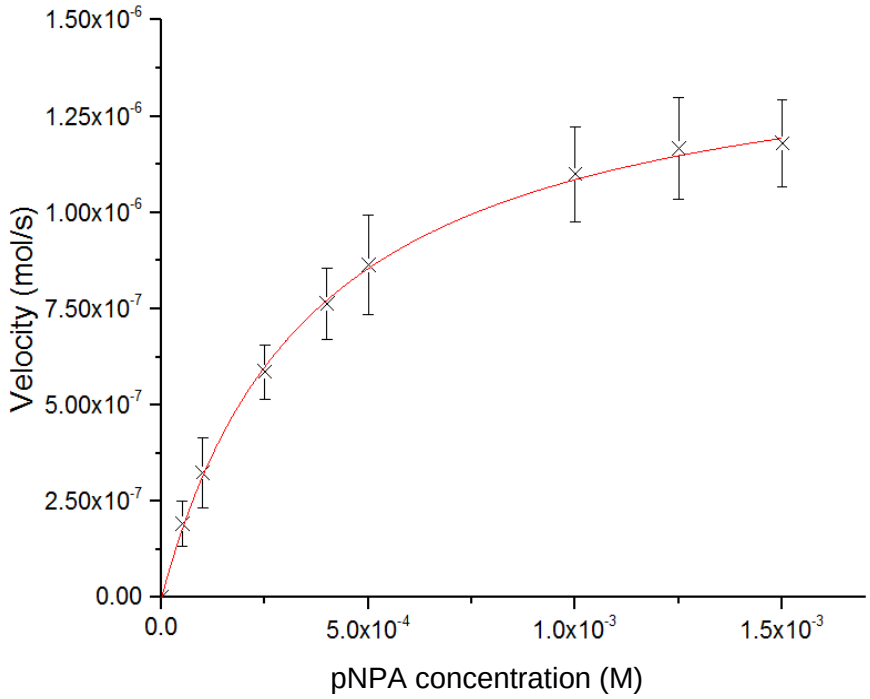


Figure 2.29: Michaelis-Menten plot for the EcBioH-C pNPA hydrolysis assay.

Kinetics parameters were determined using OriginLab software. These kinetic parameters are shown in Table 2.4 along with those previously determined by Sanishvili and colleagues.[30]

Parameter		Determined values	Reported data ([30])
Michaelis constant	K_M	0.37 ± 0.08 mM	0.29 ± 0.04 mM
Turnover number	k_{cat}	22.2 ± 2.13 s^{-1}	18.5 ± 1.7 s^{-1}
Specificity constant	k_{cat}/K_M	60.0 ± 14.2 $mM^{-1}.s^{-1}$	63.8 ± 10.6 $mM^{-1}.s^{-1}$

Table 2.4: Kinetic parameters for EcBioH pNPA hydrolysis assay.

The determined kinetic parameters were very similar to those previously determined, with the Michaelis constant and the specificity constant within the error margin, while the determined turnover number is slightly higher than the previously reported data while being of the same order of magnitude. Hence, this hydrolysis reaction data will be of significant use to compare biocatalysts esterase activity, to supplement the comparison of their MBH catalytic activities.

2.4 Optimisation of the biocatalysed MBH reaction

Before further investigating the origin of EcBioH MBH catalysis, optimised conditions had to be found. Previously, Jiang *et al.* had reported using 30 mg/mL of lyophilised lysate of *E. coli* overexpressing EcBioH to catalyse the coupling of 4-NBA with MVK with a 46% yield[20]. As the aim of this study was to use purified EcBioH to gain mechanistic insights on the biocatalysed MBH reaction, an initial conditions screen were carried. More specifically, three parameters were investigated:

- Enzyme loading
- Temperature
- Concentration of reagents

Similarly to the previous work, the conversion was measured after 96 hours.

2.4.1 Enzyme loading

The enzyme loading was the first parameter to be investigated. As Jiang's work used an unpurified lyophilised lysate, it was difficult to quantify accurately the concentration of EcBioH that was present in their reactions. As such, a few concentrations of purified EcBioH-C were tried, and the conversion was assayed after 96 hours, as showed in Table 2.5.

Enzyme loading (mg/mL)	Enzyme loading (μM)	Conversion (%)	e.e. (%)
16	542	28	Not measured
8	271	20	Not measured
4	**136**	**13**	**0**
2	68	8	Not measured
0	0	3	0

Table 2.5: Conditions screen – Enzyme loading.
Conditions: 4-NBA concentration: 4 mM; MVK concentration 8 mM; Temperature: 30ºC; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5. All reactions were run in duplicates. The conversion and e.e. were measured after 96 hours. ee was measured only for the best conditions.

Hence, from the results above, the chosen optimal purified EcBioH-C loading was 4 mg/mL, as it provides a good balance between measurable conversion and low enzyme loading. It was also found that the wild-type enzyme does not provide any enantioselectivity as shown on Figure 2.30.

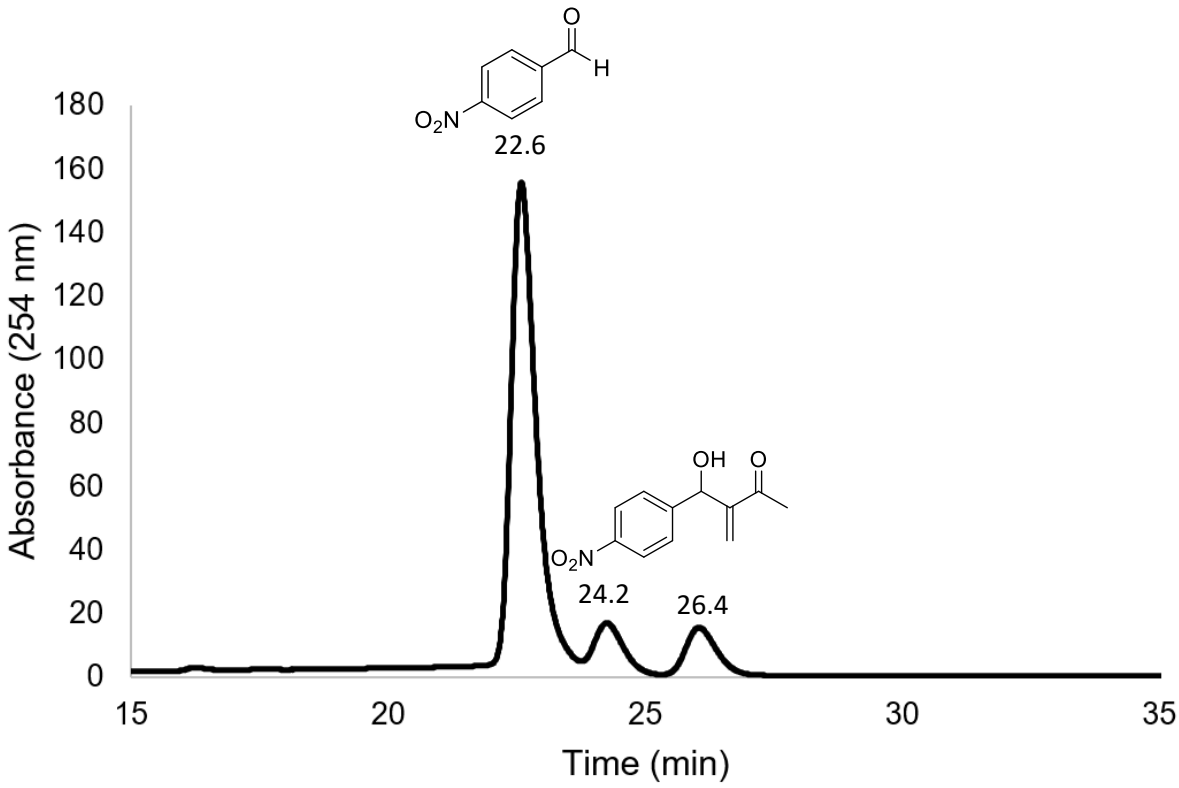


Figure 2.30: Chiral HPLC chromatogram of the biocatalysed MBH reaction. The chirality of the species corresponding to each peak is unknown.
Conditions: EcBioH-C loading: 4 mg/mL, 4-NBA concentration: 4 mM; MVK concentration 8 mM; Temperature: 30C; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5.

2.4.2 Temperature

Using 4 mg/mL EcBioH-C, the temperature was then investigated as a parameter to maximise the biocatalytic conversion. Four temperatures were investigated as showed in Table 2.6.

Temperature (ºC)	Conversion (%)
23	4
30	13
37	13
44	12

Table 2.6: Conditions screen – Temperature.
Conditions: 4-nitrobenzaldehyde concentration: 4 mM; methyl vinyl ketone concentration 8 mM; EcBioH-C loading: 4 mg/mL; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5. All reactions were run in duplicates. The conversion was measured after 96 hours.

Interestingly, it was found that although at 23ºC, very low conversion was observed, the conversion was similar at 30ºC, 37ºC and 44ºC. This suggests that this biotransformation can be carried in a wide range of conditions and that EcBioH-C is fairly thermostable. As such, the lowest temperature where significant conversion is observed, 30ºC, was chosen in the optimised conditions.

2.4.3 Concentration of reagents

The final parameter to be optimised was the concentration of reagents. First, the concentration of the limiting reagent was investigated while using one equivalent of 4-NBA and two equivalents of MVK. The results are shown in Table 2.7.

4-NBA concentration (mM)	MVK concentration (mM)	Conversion with protein (%)	Conversion without protein (%)
2	4	12	2
4	8	13	2
8	16	17	3
16	32	16	4

Table 2.7: Conditions screen – limiting reagent concentration.
Conditions: EcBioH-C loading: 4 mg/mL; Temperature: 30ºC; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5. All reactions were ran in duplicates. Conversion was measured after 96 hours.

As shown in Table 2.5, the concentration of the limiting reagent, 4-NBA does not affect the conversion dramatically. However, precipitation of the starting material was observed when using 16 mM of 4-NBA. As such, 8 mM of 4-NBA were used in this study. Then the number of equivalents of MVK was optimised as shown in Table 2.8. HPLC chromatograms for the EcBioH-C-catalysed reactions with each number of equivalents are shown on Figure 2.31.

Number of equivalents of MVK	MVK concentration (mM)	Conversion with protein (%)	Conversion without protein (%)
1	8	4	1
2	16	17	3
4	32	37	5
8	64	55	9
16	128	62	16

Table 2.8: Condition screen – number of equivalents of methyl vinyl ketone.
Conditions: 4-NBA concentration: 8 mM; EcBioH-C loading: 4 mg/mL; Temperature: 30C; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5. All reactions were run in duplicates. The conversion was measured after 96 hours.

Changing the number of equivalents of MVK has a significant effect on the conversion. It was found that increasing the number of equivalents from 2 to 8, leads an increase in the conversion from 17% to 55%. This increase could be explained by the fact that adding more MVK should drive the reaction equilibrium towards the formation of the MBH product. However, it was found that 8 equivalents of MVK were the optimal loading. Indeed, adding 16 equivalents of MVK only provides an improvement of the enzyme-free conversion. It was postulated that this was because, as MVK is very reactive, large concentrations of this compound would inhibit the reaction. This suggestion was strengthened by the fact that slow and partial precipitation of EcBioH-C was observed at all concentrations of the MVK substrate.

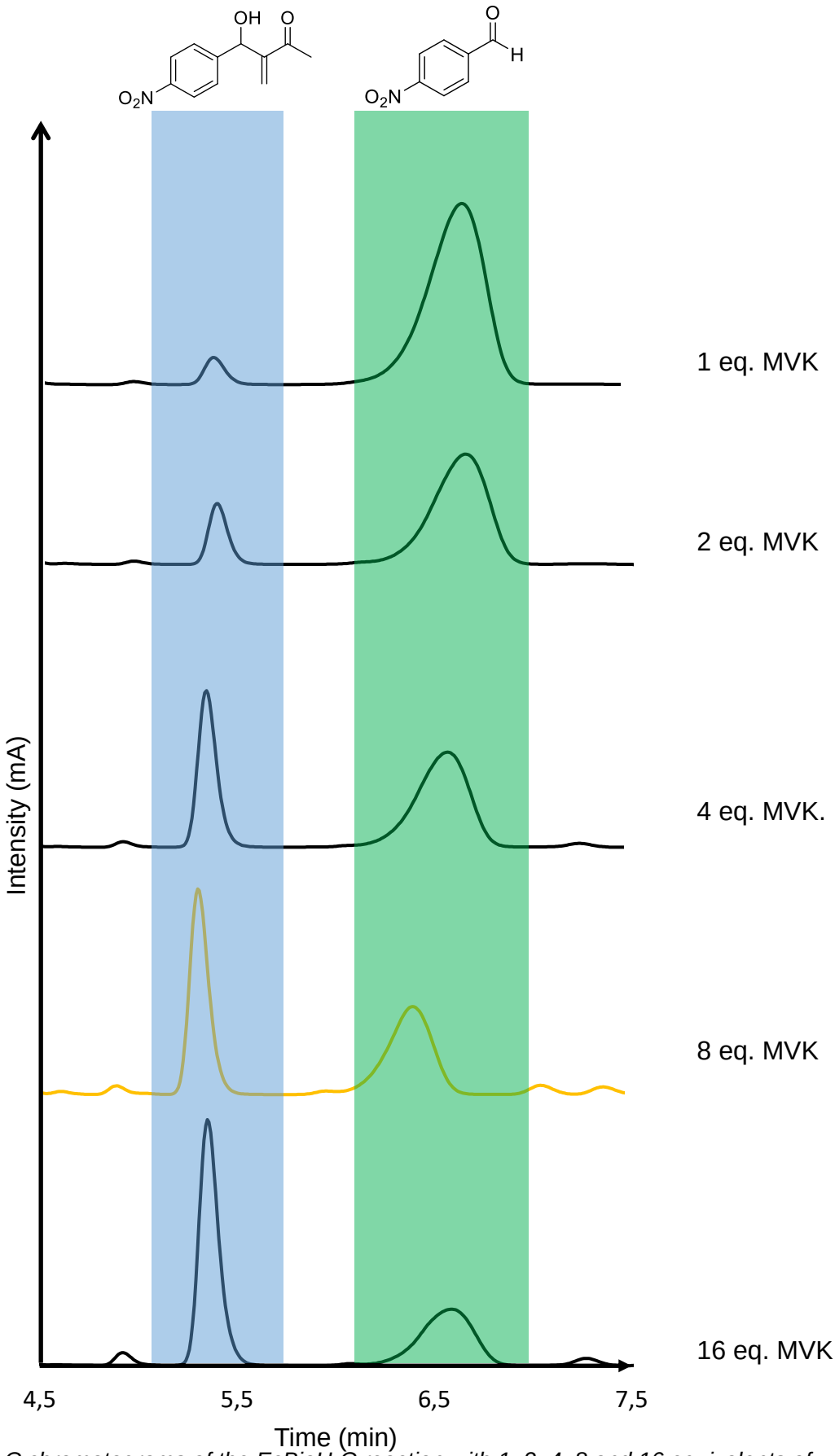


Figure 2.31: HPLC chromatograms of the EcBioH-C reaction with 1, 2, 4, 8 and 16 equivalents of MVK after 96 hours.
Conditions: 4-NBA concentration: 8 mM; EcBioH-C loading: 4 mg/mL; Temperature: 30C; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5.

2.5 Investigating the involvement of the catalytic triad

As the main aim of this work is to study the mechanism of the biocatalytic MBH reaction, a lot of effort was dedicated to the identification of the MBH catalytic active site. Indeed, as discussed in the introduction, there is a disagreement on the origin of the MBH catalytic activity and, more specifically, on the identity of the nucleophilic residue that undergoes the nucleophilic attack on the electron-withdrawing alkene, in the first step (I) of the MBH mechanism shown on Figure 2.32.

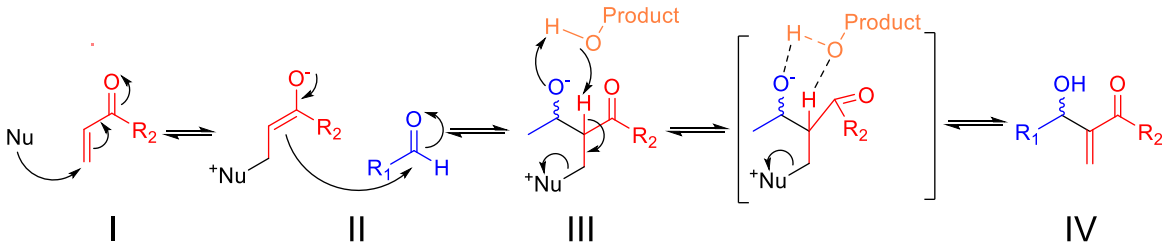


Figure 2.32: MBH reaction proposed mechanism for a vinyl ketone, an example of an electron-withdrawing alkene. The nucleophilic catalyst undergoes Micheal addition on the alkene (I), forming an extremely nucleophilic zwitterion which will attack the electrophilic aldehyde (II). A product-catalysed proton transfer (III) and elimination (IV) of the catalyst afford the product.

Indeed, it has been postulated that this nucleophilic catalyst could be the side chain of a cysteine[17], a histidine[18], a lysine[16] or a serine that is found in a catalytic triad[16], while another hypothesis states that there is general protein catalysis where any histidine residue could trigger the reaction[37]. Hence some of these postulated hypotheses were tested to identify that nucleophilic residue.

First, the serine that is part of the catalytic triad was investigated. As previously mentioned, EcBioH has two postulated catalytic triads, the "classic" one, composed of Ser82, Asp207 and His235, shown on Figure 1.15, is more likely to be of catalytic importance for the MBH reaction, since it was found that this deactivation of these residues lead to loss of esterase activity. The "cryptic" triad showed in Figure 1.16 is made of Ser53, Asp46 and His20. Upon inspecting the available BioH crystal structures, it seems that this triad is not involved in catalysis as the residues of this triad are not within H-bond distance from each other. However, as these residues are located on very flexible loops, the residues may swing to shorter distance and catalyse the reaction . This could be determined through structural studies of the protein-ligand complex. For completeness, both catalytic triads will be deactivated to investigate their roles in MBH biocatalysis.

2.5.1 Generation of the mutants

The EcBioH-C mutants were generated using the overlapping primers site-directed mutagenesis technique developed by Liu and Naismith[38]. In this method, the pair of primers used to introduce the mutation are constituted of two parts:

- A complementary region, where the forward and reverse primers are complementary to each other, which contains the desired mutation.

- A non-overlapping region, significantly larger than the complementary region, which is complementary to the template DNA.

All the mutations were introduced successfully and the mutated DNA was used to transform and express *E. coli* BL21 (DE3) cells and the mutant was purified similarly to the wild-type enzyme. Mutants containing S82A, H235A and S55A mutations gave similar yields to the wild-type enzyme (10+ mg/L of culture). However, the D207A mutant resulted in very poor yields (approx. 0.8 mg/L of culture).

SDS-PAGE analysis (Figure 2.33) demonstrated that the poor yield results from the fact that this mutation leads to the insolubility of the protein. Indeed, substantial changes in protein solubility due to the addition or removal of an aspartic acid were already described in the literature[39]. Aspartic acid is very hydrophilic because the free carboxylic acid on its side chain will be deprotonated at biological pH as its pKa is 2.68[40]. Hence, the aspartic acid will be a potent H-bond acceptor both useful as part of the catalytic triad, as developed earlier, but also valuable for creating key interactions with water molecules to solubilize the protein.

Because of this poor solubility, the EcBioH-C D207A mutant was only purified by nickel affinity chromatography as it was found by SDS-PAGE that this purification step was sufficient to obtain the pure protein. The protein was then buffer exchanged into the MBH reaction (gel filtration) buffer by diafiltration using a 10 kDa spin column. Furthermore, because of these solubility issues, the full catalytic triad knockout mutant (S82A D207A H235A) was not generated.

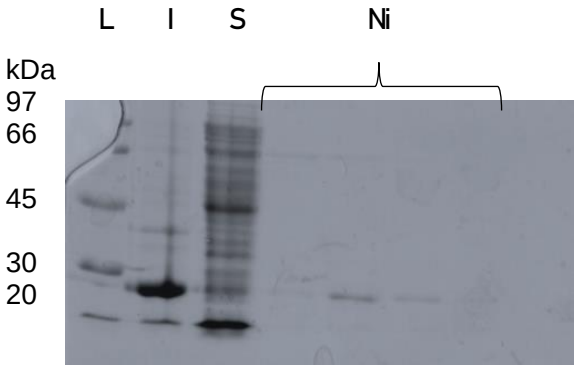


Figure 2.33: 12% SDS-PAGE gels of a hexahistidine-tagged EcBioH-C D207A purification.
L: Low molecular weight ladder; I: Insoluble fraction (cell pellets); S: Soluble fraction (cell pellets); Ni: After immobilised nickel affinity chromatography;
A large band at EcBioH-C expected mass (26 kDa) is observed in the insoluble fraction.

All the purified mutants (S82A; S82A D207A; S82A H235A; S55A) were characterised by ESI-MS. The results are shown in Table 2.9

Phenotype	Theoretical molecular weight (Da)	Determined molecular weight (Da)
S82A	29554.13	29553.20 ± 1.01
S82A D207A	29510.12	29514.17 ± 0.59
S82A H235A	29488.07	29487.88 ± 0.68
S53A	29554.13	29553.43 ± 0.76

Table 2.9: MS data of the EcBioH-C mutants.
** Theoretical molecular weights were calculated by ExPASy-ProtParam tool. Average mass is shown.*

2.5.2 Assaying the mutants

The generated mutants were assayed for their MBH and esterase activities. The MBH activity was determined from the HPLC assay presented in Section 2.2.1 while the esterase activity was measured using the pNPA hydrolysis assay showed in Section 2.3. The results for these assays are presented in Table 2.10.

Phenotype	MBH conversion (%)	Fold improvement over parent	Esterase turnover number (s^{-1})	Fold improvement over parent
WT	55 ± 2	1.000	22.2 ± 2.13	1.000
S82A	58 ± 3	1.054	0.38 ± 0.02	0.017
S82A D207A	53 ± 2	0.963	0.42 ± 0.04	0.019
S82A H235A	57 ± 3	1.036	0.13 ± 0.01	0.006
S53A	55 ± 3	1.000	15.8 ± 2.58	0.712

Table 2.10: MBH and esterase assay results for EcBioH-C WT and a range of EcBioH-C mutants with deactivated catalytic triads residues.
Conditions: 4-NBA concentration: 8 mM; MVK loading: 64 mM; EcBioH-C loading: 4 mg/mL; Temperature: 30C; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5.

Hence, while the esterase catalytic efficiency drops significantly for all the main catalytic triad knockout mutants with an almost 100-fold decrease in activity, it was interesting to observe that the MBH conversion remains similar for all these mutants. Indeed, all the results are contained within the wild-type error. This suggests that the main catalytic triad does not have an effect in MBH biocatalysis.

Similarly, no change in MBH activity was observed when deactivating the second cryptic catalytic triad (S53A mutant). This suggests that this catalytic triad is not involved in MBH catalysis, hence not supporting the hypothesis that a serine part of a catalytic triad is responsible for MBH catalysis. Interestingly, however, a significant decrease in esterase activity was observed upon deactivating the auxiliary catalytic triad. While, at first, this could suggest that about 30% of the esterase activity originates from this auxiliary triad, this hypothesis is not supported by the fact that upon deactivating the main catalytic triad, a 98-99.5% loss of activity is observed. This would mean that the auxiliary triad contributes to up to 0.5% of the esterase activity. However, it is postulated that deactivating the auxiliary catalytic triad might disrupt the folding of the enzyme, resulting in a less efficient arrangement of the catalytic triad or a less easily accessible active site which could explain the observed 30% loss in activity.

2.5.3 Mass spectrometry analysis of the EcBioH covalent product

To further confirm the non-involvement of the catalytic triad, mass spectrometry experiments were also performed. The aim of these experiments was to verify that, when the catalytic serine is modified, the first step of the MBH reaction (the Micheal addition of the biocatalyst, Figure 2.34) still occurs and that a protein-cyclohexenone covalent adduct is still formed.

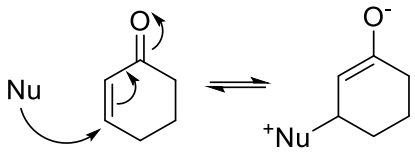


Figure 2.34: The proposed first step of the biocatalysed MBH reaction is the Micheal addition of a nucleophilic residue on cyclohexanone.

Cyclohexenone (96 Da) was used for this study instead of MVK (70 Da) as it provides a larger, more pronounced mass difference. The MBH coupling of 4-nitrobenzaldehyde with cyclohexenone was found to be almost as efficient as MVK with a conversion of 10% after 96 hours, using non-optimised conditions, which resulted in 13% conversion for the coupling of 4-NBA and MVK. Hence, EcBioH-C WT and EcBioH-C S82A were incubated with phenylmethylsulfonyl fluoride (PMSF) (1 mM). This reagent, which serves as a serine protease inhibitor[41], binds covalently modifies the catalytic serine and deactivates it. This reaction is a simple nucleophilic substitution where PMSF's fluorine atom is substituted by the catalytic serine, forming a protein-phenylmethylsulfonyl (PMS) covalent adduct. The mechanism is shown in Figure 2.35.

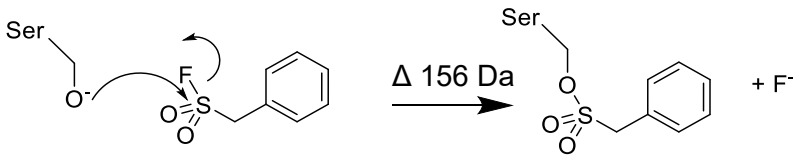


Figure 2.35: Mechanism of the PMSF deactivation of the catalytic serine.

After one hour incubation, the deactivation of the EcBioH-C WT catalytic serine by PMSF was observed by ESI-MS. In contrast, the EcBioH-C S82A mutant, where the catalytic serine was mutated, did not react with PMSF. Cyclohexenone (1 mM) was subsequently added and the reaction was incubated for one more hour as shown on Figure 2.36.

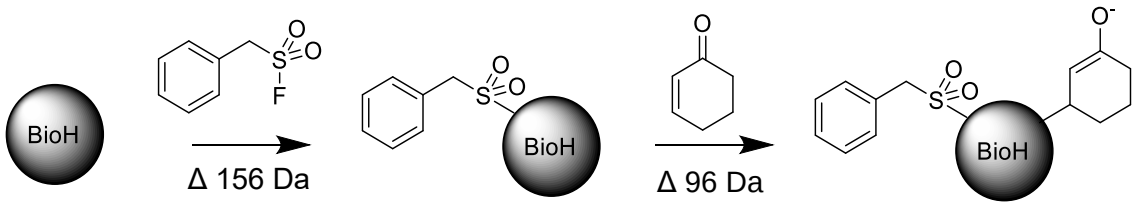


Figure 2.36: ESI-MS experiment scheme. Both PMSF and cyclohexanone were added to EcBioH-C, leading to an increase in mass which was monitored by LC-ESI-MS.

As expected, the complex EcBioH-C WT + PMS + cyclohexanone (+252 Da) was observed by ESI-MS (Figure 2.37), providing further support of the non-involvement of the catalytic triad in the MBH reaction catalysis.

Bottom-up proteomics was attempted and the sample was digested using trypsin to generate protein fragments which were analysed by Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS). However, these analyses were unsuccessful due to poor coverage. Hence this technique could not provide further confirmation that the PMS was bound to Ser82.

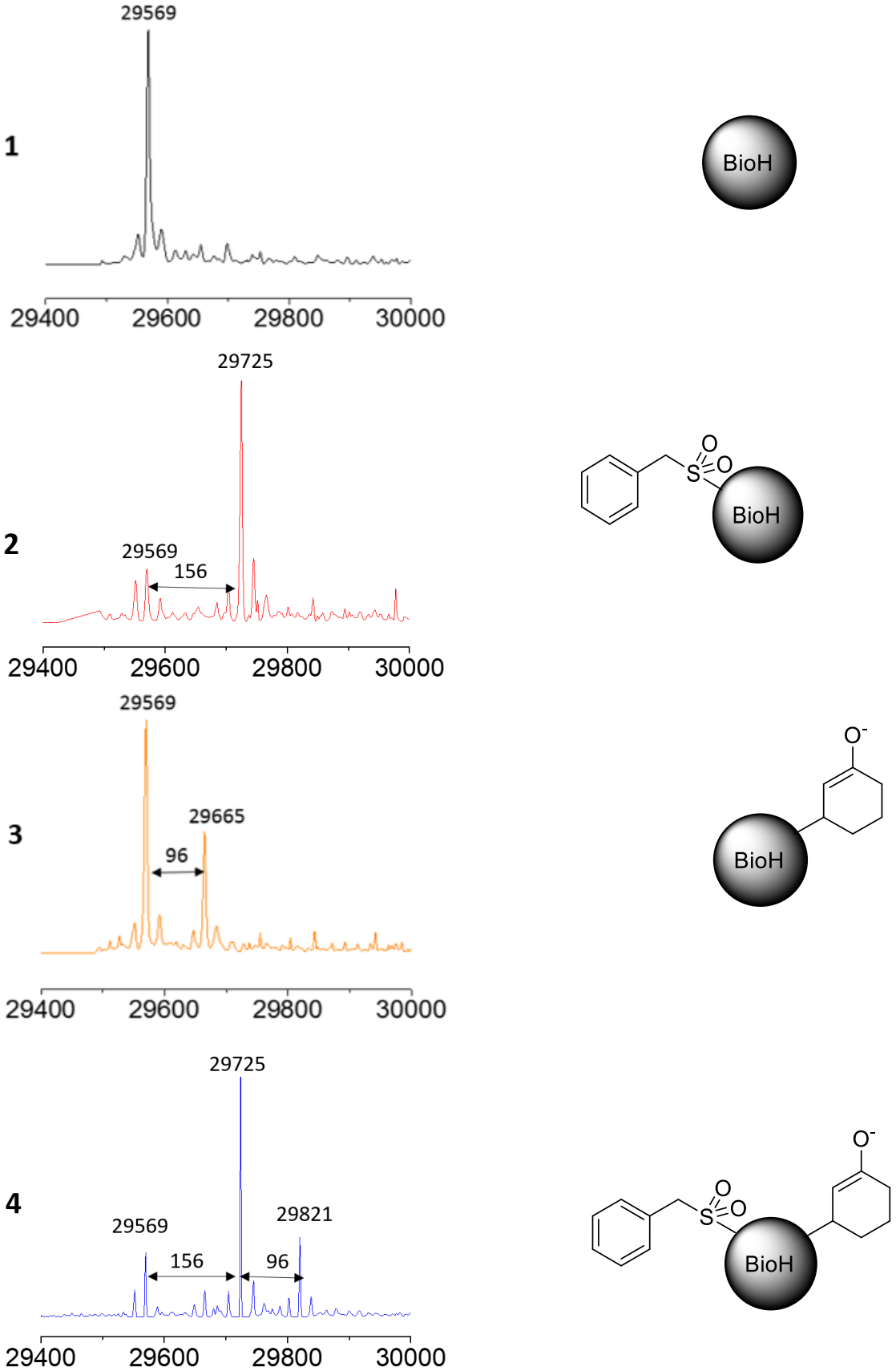


Figure 2.37: 1: Deconvoluted MS(ESI) spectrum for: 1: BioH only – only one peak is observed at 29569 Da.
2: BioH + PMSF – 2 peaks are observed: one small corresponding to unbound BioH and one large peak at +126 Da corresponding to BioH + PMS.
3: BioH + cyclohexenone – 2 peaks are observed: one corresponding to unbound BioH and one smaller one at +96 Da corresponding to the BioH + cyclohexanone complex.
4: BioH + PMSF + cyclohexenone – 3 peaks are observed: one corresponding to BioH only, one at +126 Da corresponding to BioH + PMS and one at + 222 Da corresponding to BioH + PMS + cyclohexanone.

2.6 Investigating the involvement of cysteine

2.6.1 Cysteine, another potent nucleophile

Another potential nucleophile which could trigger the first Michael addition in the MBH reaction is cysteine (pKa = 8.3)[40]. At the reaction pH (7.5) the thiol should be reduced, unless the microenvironment of the active site sufficiently alters the pKa of the cysteine residue to allow it to be oxidised[42]. Indeed, while being one of the least abundant amino acids, an analysis of the distribution of conservation levels of residues within enzymes showed that cysteine is one of the most conserved residues, with a conservation degree of 90+%[43]. Examples of the nucleophilicity of cysteines include the whole class of cysteine protease enzymes. These enzymes share a common mechanism which involves a nucleophilic cysteine close to a basic residue such as a histidine in a catalytic dyad. These enzymes degrade proteins through cleaving their peptide bonds and are commonly encountered in fruits such as fig or pineapple[44]. Their proposed mechanism of action is shown in Figure 2.38[45].

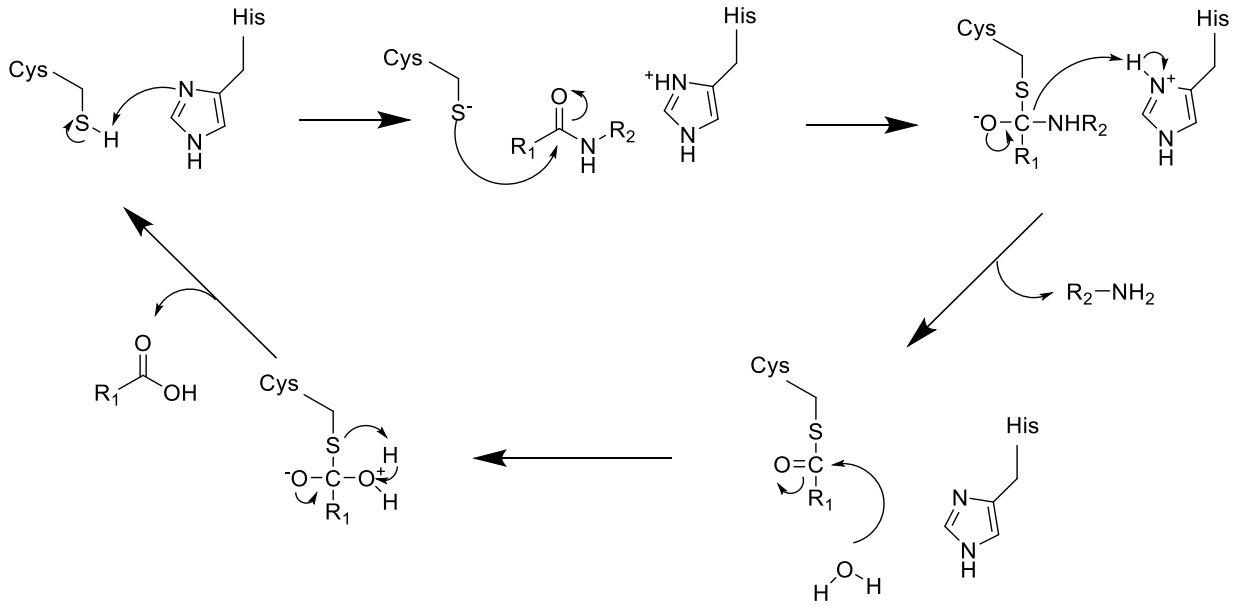


Figure 2.38 Cysteine protease reaction mechanism: Upon deprotonation by histidine, cysteine will undergo nucleophilic substitution on the peptide to be cleaved. The amine (C-ter end of the peptide) will be released first. Then an hydrolysis reaction will occur to free the acid (N-ter end of the peptide).

2.6.2 Cysteine alkylation experiments

To investigate the involvement of the free thiol contained in cysteine residues, reactive cysteines should be alkylated to examine if the MBH reaction still occurs. It was found that iodoacetamide (IAA) is a useful reagent since it undergoes selective alkylation of reactive cysteine residue(s)[41]. The cysteine would undergo nucleophilic substitution on the iodoacetamide, generating a covalent, acetamide

adduct[53]. The protein was then assayed for the MBH reaction. The mechanism of this alkylation is shown in Figure 2.39 and the results of this assay are shown on Table 2.11.

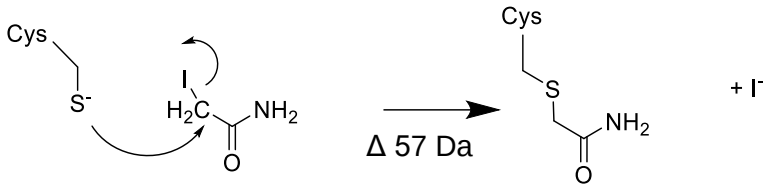


Figure 2.39: Mechanism for the active cysteine alkylation by iodoacetamide.

Reactive cysteine residue(s)	MBH conversion (%)
Not alkylated (Control)	55 ± 2
Alkylated (with iodoacetamide)	56 ± 4

Table 2.11: Results of the MBH assay with and without alkylating the reactive cysteine(s). Conditions: 4-NBA concentration: 8 mM; MVK loading: 64 mM; EcBioH-C loading: 4 mg/mL; Temperature: 30C; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5.

Following this assay, it was demonstrated that alkylating cysteine residues does not affect the MBH conversion, suggesting that cysteine does not act as a nucleophilic activator for the biocatalysed MBH reaction. Alkylation of the cysteine was proven by LC-MS (ESi). To give further evidence of the non-involvement of this cysteine, MS-adducts experiments were undergone, similarly to Part 2.5.3 using IAA to alkylate any reactive cysteine residues. Hence, at first EcBioH-C WT (0.1 mM) was incubated for one hour with IAA (1 mM). Full alkylation of one cysteine was observed by LC-MS(ESI) as one single peak corresponding to the mass of EcBioH-C + one acetamide was observed. This suggests that, out of the three cysteine residues contained in EcBioH protein sequence, only one is accessible. Then the alkylated EcBioH-C was incubated for one hour with cyclohexenone (1 mM) and the EcBioH-C WT + acetamide + cyclohexenone was detected by LC-MS(ESI). The reaction scheme is shown on Figure 2.40 and the MS spectra obtained are shown on Figure 2.41. The fact that cyclohexenone binds EcBioH-C when the reactive cysteine is alkylated provides further evidence on the non-involvement of a reactive cysteine residue in MBH catalysis.

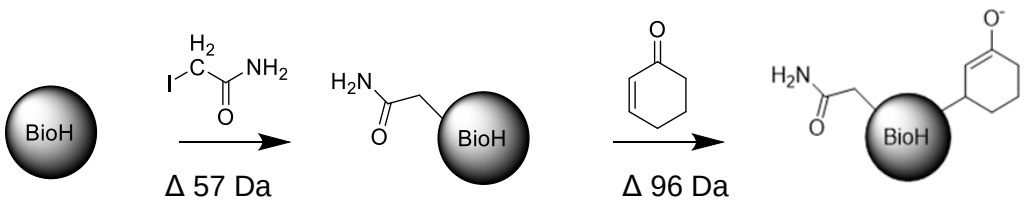


Figure 2.40: ESI-MS experiment scheme. Both PMSF and cyclohexanone were added to EcBioH-C, leading to an increase in mass which was monitored by LC-ESI-MS.

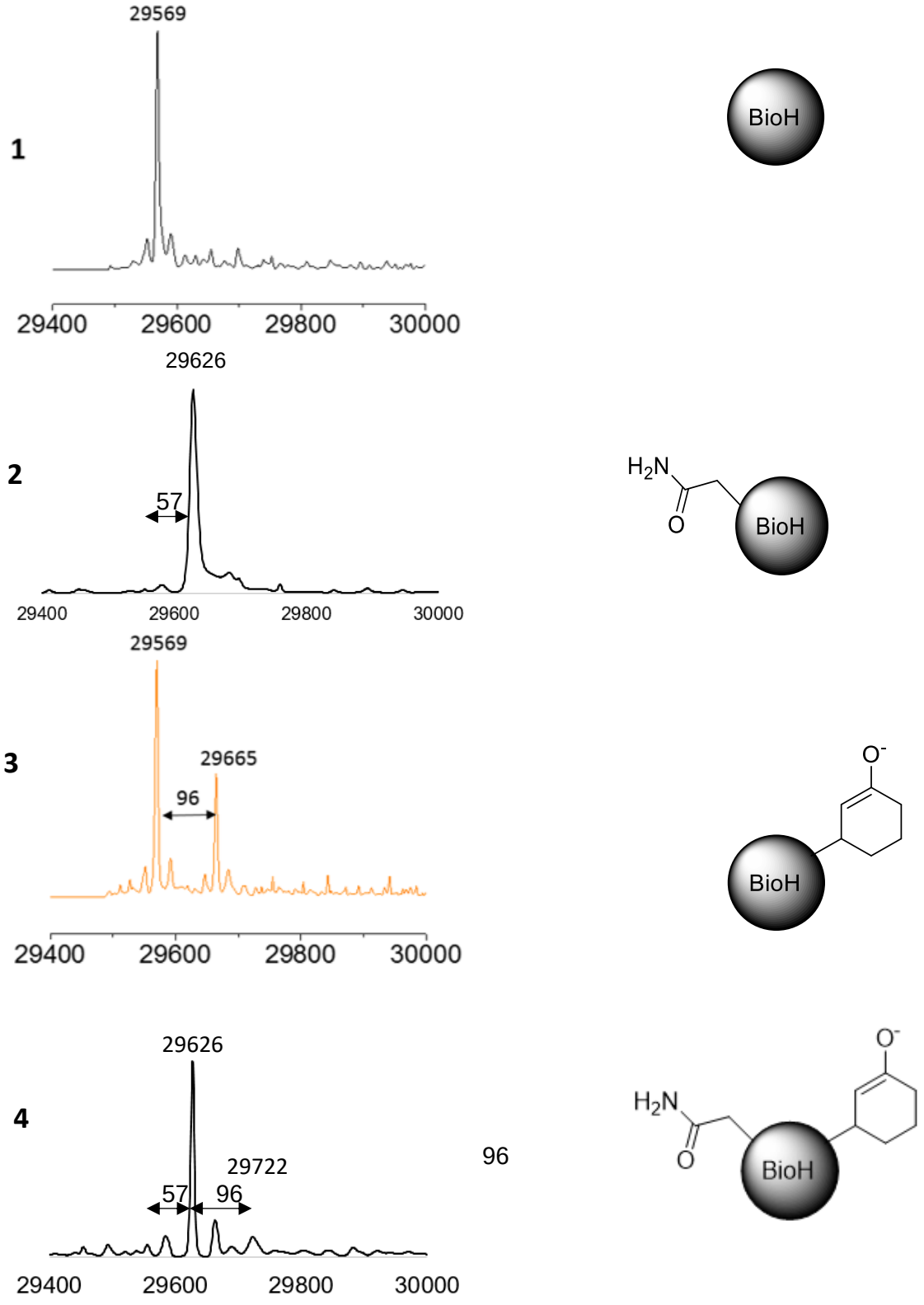


Figure 2.41: 1: Deconvoluted MS(ESI) spectrum for: 1: BioH only – one peak is observed at 29569 Da.
2: BioH + IAA – One peak at +57 Da corresponding to BioH + acetamide.
3: BioH + cyclohexenone – 2 peaks are observed: one corresponding to unbound BioH and one smaller one at +96 Da corresponding to the BioH + cyclohexanone complex.
4: BioH + IAA + cyclohexenone – 2 peaks are observed: one at BioH + 57 Da corresponding to BioH + acetamide and one at + 153 Da corresponding to BioH + acetamide + cyclohexanone.

Bottom-up proteomics was attempted and the sample was digested using trypsin to generate protein fragments which were also analysed by FTICR-MS. However, these analyses were unsuccessful due to poor coverage. Hence this technique could not provide insights on which cysteine is targeted by IAA.

2.7.1 Investigating another pimeloyl-ACP methyl ester esterase

The involvement of the ABHF and its role in the BioH-catalysed MBH reaction was then investigated. The rationale behind this is, while the approach explored so far was that the first step of the catalysis was the activation of the electrophile by a highly nucleophilic residue, another effect to consider is the stabilization of the oxyanions intermediates by the oxyanionic hole of the enzyme. Such H-bond interaction could favour the reaction from delocating some electron density from the oxyanions.

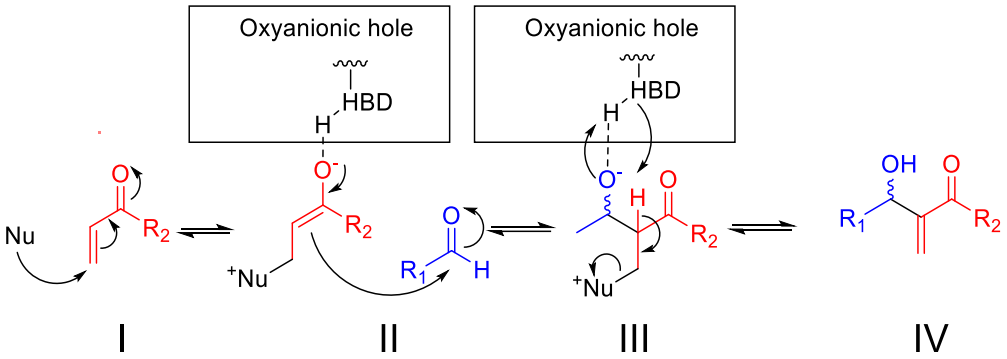


Figure 2.42: Second proposed MBH biocatalysed reaction mechanism: As the partial positive charge of the vinylic carbon will be increased because of the carbonyl H-bond interaction, weaker nucleophiles might be able to undergo Micheal addition.

Carbonyl activation through H-bonding interaction is a well-documented mechanism triggering various reactions, including the Diels-Alder[46-47] reaction and the aldol reaction[48].

2.7.2 BioG, an atypically folded enzyme

Hence, to assess the involvement of the enzyme fold, another pimeloyl-ACP methyl ester esterase, BioG from *H.influenzae* (HiBioG) was used. As discussed in section 2.1.4, the structure of HiBioG, which has 25% identity to EcBioH was determined in 2016 by the group of John Cronan. It revealed that HiBioG possesses an atypical α/β-hydrolase fold[34]. While on BioH, and typical ABHF enzymes, the parallel β-sheets are circled by α-helices on both sides, large loops replace four α-helices in HiBioG. This feature could be therefore giving HiBioG more spacial flexibility than EcBioH. As a result this

enzyme might display more promiscuity[34]. The structure of HiBioG, showing that loops have replaced α-helices of a typical α/β hydrolase fold, is shown in Figure 2.43.

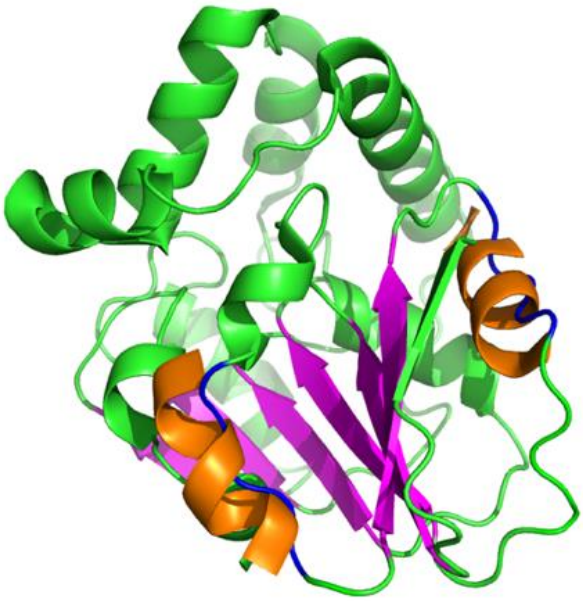


Figure 2.43: HiBioG crystallographic structure highlighting the atypical α/β hydrolase structure (PDB ID: 5GNG[34]). The six parallel β-sheets are shown in magenta. In blue are shown BioG atypical loops to which are superimposed in orange the EcBioH (PDB ID: 1M33) α-helices expected to be observed on an α/β hydrolase fold.

HiBioG pET28 plasmid was obtained from Prof John Cronan's lab (University of Illinois Urbana-Champaign). The N-terminal histagged enzyme was expressed and purified following the published procedure[34]. Yields of 12 mg/L of culture, comparable to the ones obtained for EcBioH, were obtained. The SDS-PAGE gel, along with the LC-ESI-MS spectrum, are shown in Figure 2.44.

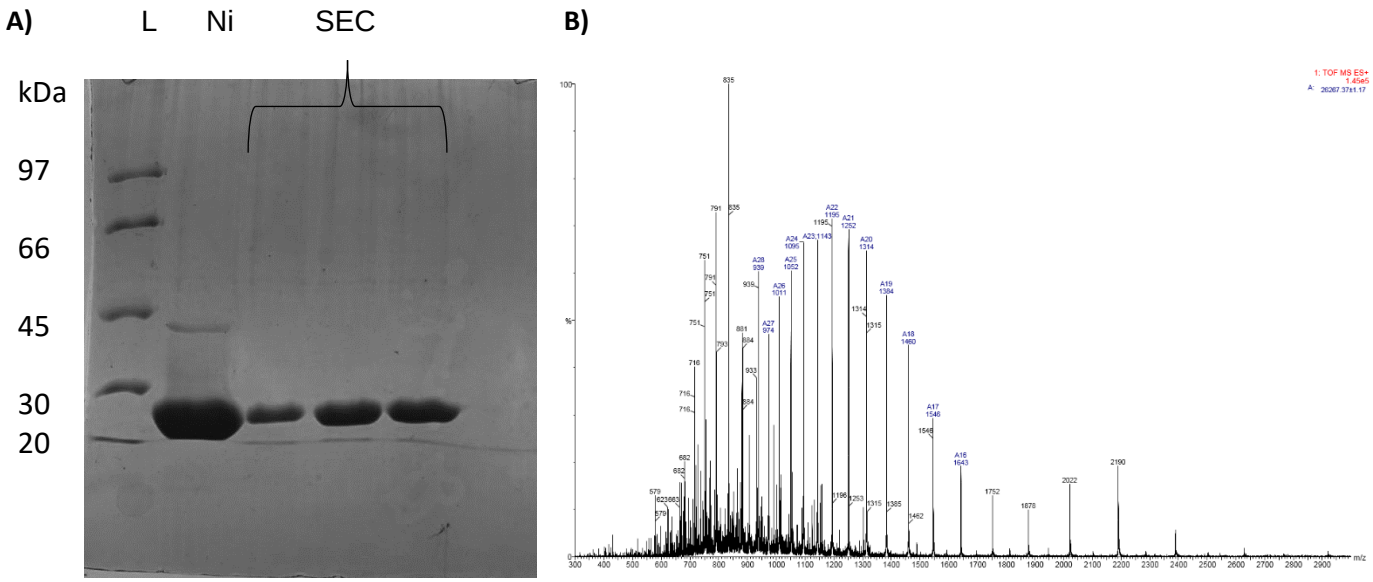


Figure 2.44: A) 12% SDS-PAGE gel for the purification of HiBioG. L: Low molecular weight ladder; Ni: Sample post-Ni affinity chromatography; SEC: Sample post-size-exclusion chromatography. B) LC-ESI-MS spectrum of HiBioG. Theoretical molecular weight: 26266.73 Da, Experimental molecular weight: 26267.37 ± 1.17 Da.

2.7.3 Assaying HiBioG

HiBioG was assayed both for their pNPA esterase and MBH activities. The Michealis-Menten plot for HiBioG pNPA esterase activity is shown on Figure 2.45 and the results of the pNPA esterase and the MBH assays are shown on Table 2.12.

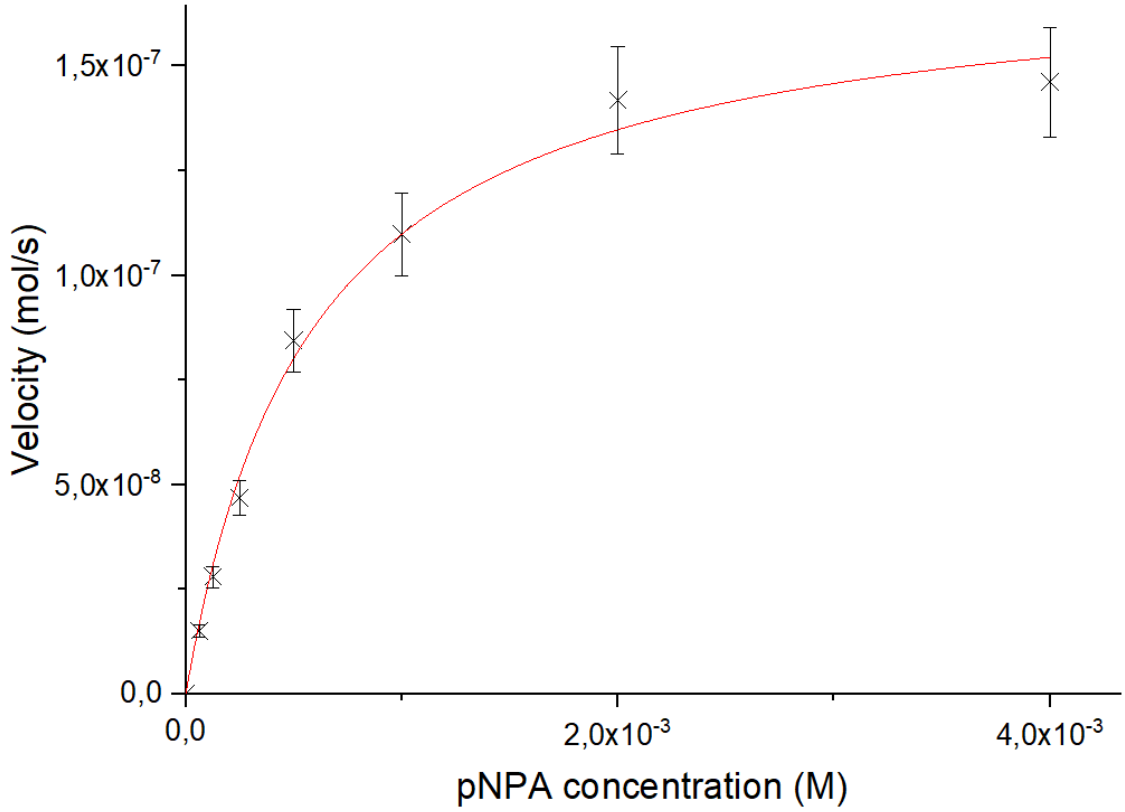


Figure 2.45: Michaelis-Menten plot for the HiBioG pNPA hydrolysis assay.

Parameter		HiBioG values	EcBioH values
Michaelis constant	K_M	0.59 ± 0.06 mM	0.37 ± 0.08 mM
Turnover number	k_{cat}	2.25 ± 0.12 s^{-1}	22.2 ± 2.13 s^{-1}
Specificity constant	k_{cat}/K_M	5.51 ± 0.60 $mM^{-1}.s^{-1}$	60.0 ± 14.2 $mM^{-1}.s^{-1}$
MBH activity		55 ± 2	58 ± 3

Table 2.12: pNPA esterase and MBH assay results for and HiBioG WT and EcBioH-C WT.

Hence, it was interesting to find that HiBioG also displays MBH activity and affords similar conversion to EcBioH for the MBH coupling of 4-NBA and MVK. These results suggest that ABHF structure does not affect the MBH reaction significantly. Interestingly, it was also found that, while HiBioG and EcBioH have similar Michealis constants, suggesting that pNPA have similar binding affinities for both enzymes, HiBioG turnover number is almost 10 times smaller than for EcBioH. It can therefore be concluded than HiBioG is much slower than EcBioH. This suggests that the increased flexibility given by the loops has a negative impact on the esterase activity.

2.8 Assaying the involvement of histidine in the EcBioH-catalysed MBH reaction

2.8.1 Histidine, a weak catalyst for the chemical MBH reaction

Previously, nucleophilic bases have been found by Morita, Baylis and Hillman to catalyse the C-C bond coupling of an aldehyde with an electron-poor alkene. While tricyclohexylphosphine and DABCO were first exemplified as suitable MBH reaction initiators, most research on the MBH reaction aimed at making enantioselective catalysts. However, simpler bases were also investigated by Luo *et al.* [49]. They have found that the coupling of cyclopent-2-enone with 4-NBA could be "catalysed" by one equivalent of imidazole with a 92% yield as shown on Figure 2.46[49]. Imidazole appears to be a better base for this coupling than DABCO, which only produced a 32% yield[49]. This result was somewhat surprising as DABCO was found to be much more nucleophilic than imidazole[50]. The same group subsequently showed that one equivalent of L-histidine in a weakly basic aqueous solution (1M sodium bicarbonate) could also catalyse the coupling shown on Figure 2.46 with a 65% yield[51]. However, all these results were obtained using very high concentrations of imidazole/histidine (200-250 mM).

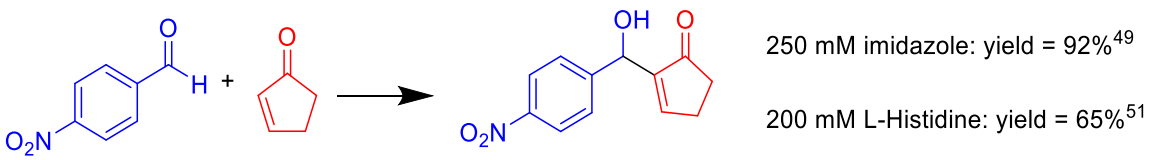


Figure 2.46: imidazole- and L-histidine-catalysed MBH reaction schemes

Some attempts to understand the mechanism of the imidazole/histidine catalyzed MBH reaction were made. Interestingly, Pereira *et al.*, in their work aiming at creating a MBH ionic liquid system, postulate that the histidine free amine (which would be part of the backbone in a protein scaffold) attacks the vinyl ketone derivative[52]. This hypothesis was formulated after they observed a slight increase of yield when they replaced imidazole by histidine and is hard to believe as this amine is protonated at pHs below 9.2 and that there is no mention of highly basic pH[52] (the first intermediate of this proposed mechanism is shown on Figure 2.47).

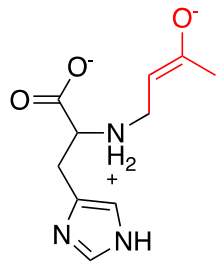


Figure 2.47: Structure of the first MBH intermediate, as postulated by Pereira et al.[52]

They postulated this after undertaking NMR and IR experiments which, unfortunately, were not discussed or included in the paper. Indeed, only MS experiments are presented in length, but these experiments are unable to determine which "part" of the histidine molecule attacks the electron-poor alkene.

An alternative mechanism is the attack of the vinyl ketone derivative by the from the imidazole "end" of the molecule and the stabilisation of the oxyanionic intermediates thanks to an H-bond between the oxyanion and a proton of the positively charged free amine, hence explaining the slight activity increase upon substituting imidazole with histidine as shown on Figure 2.48.

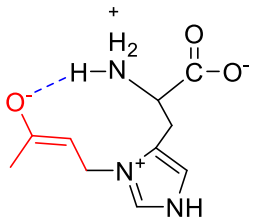


Figure 2.48: Proposed structure of the first intermediate of the histidine-catalysed MBH reaction. The oxyanion in the first step is stabilised by an H-bond interaction with the protonated amine (vinyl ketone derivative showed in red, H-bond interaction in blue).

As already mentioned, catalysts containing both a Lewis basic site which is able to attack the electron-poor alkene and also a Lewis acidic site which can stabilise the first and second oxyanions are very common for the MBH reaction. Raich *et al.* have described such catalyst where the Lewis basic site, which attacks the electron-poor alkene is an imidazole derivative[53]. Indeed, they have found that bicyclic imidazolyl alcohol (BIA) derivatives provide an improvement in activity compared to imidazole, while they have similar nucleophilicities[53]. Through computational approaches, they were able to show that, while a significant part of that improvement is due to the fact that they facilitate the proton transfer observed in the third step of the MBH reaction, part of this improvement is also caused by the fact that the oxyanionic intermediates will be stabilised by an H-bond network between the oxyanion and the hydroxyl group of the BIA derivative, *via* a water molecule, as shown on Figure 2.49.

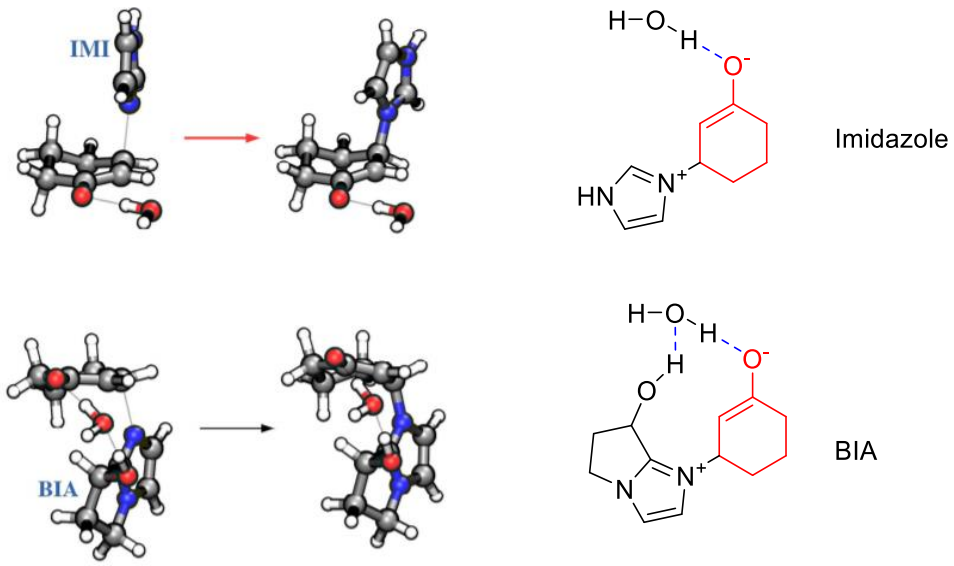


Figure 2.49: Proposed structures of the first oxyanionic intermediates in the cyclohexenone MBH reaction catalysed by imidazole (IMI - top) or by a bicyclic imidazolyl alcohol (BIA - bottom). An H-bond network is stabilising the oxyanionic intermediate for the BIA-catalysed reaction. Adapted from [28].

2.8.2 Proposed role of acid/base catalysis

With a proposed role for imidazole, albeit at 200-250 mM concentrations, the role of the hexahistidine tag was to be explored. For this reason, it is important to control at this stage that the activity of the enzyme is not (solely) due to the basicity of histidine. There are 15 histidine residues in the wild type clone of EcBioH-C used so far for this study (9 in EcBioH sequence and 6 in the hexahistidine tag). EcBioH-C concentrations of 136 µM are used in this study. This means that the effective concentration of histidine residues is 15 x 136 = 2040 µM. As it is very unlikely that all of them are accessible, this should be an overestimation. So the MBH reaction was set up using the same conditions than the biocatalysed one, except that, instead of using an EcBioH-C, 2000 µM L-histidine was used. The results are shown in Table 2.13:

Reagent	Concentration (mM)	Conversion after 96 hours (%)
L-histidine	2	16 ± 1
EcBioH-C	0.13	55 ± 2
Buffer only	N/A	9 ± 2

Table 2.13: MBH conversion using L-histidine.

Hence, using concentrations of histidine which are in line with the anticipated concentrations in the biocatalytic assay results in much lower conversion than for the enzyme is observed. Hence it can be concluded that EcBioH-C activity doesn't result solely from the basicity of the histidine residues.

2.8.3 Investigating the involvement of the hexahistidine tag

As described earlier, it was previously shown that the MBH reaction could be sharply improved if both a nucleophilic Lewis base and a Lewis acid, to stabilise the oxyanion, are present. It could be hypothesised that the hexahistidine tag which was used for this project for easy protein purification can be involved. It is thought that histidine can undergo the Micheal addition of the electron-withdrawing alkene, as showed by Luo *et al.*[24,26]. Then, the formed oxyanion could be stabilised by a hydrogen-bond with neighbouring histidine of the tag, either directly or mediated by a water molecule as described by Raich and colleagues[28]. In turn, the neighbouring histidine will act as H-bond donor which will increase its nucleophilicity and favour its Micheal addition on another MVK molecule, as shown on Figure 2.50. The hexahistidine tag could also ease the proton transfer of the third step of the reaction.

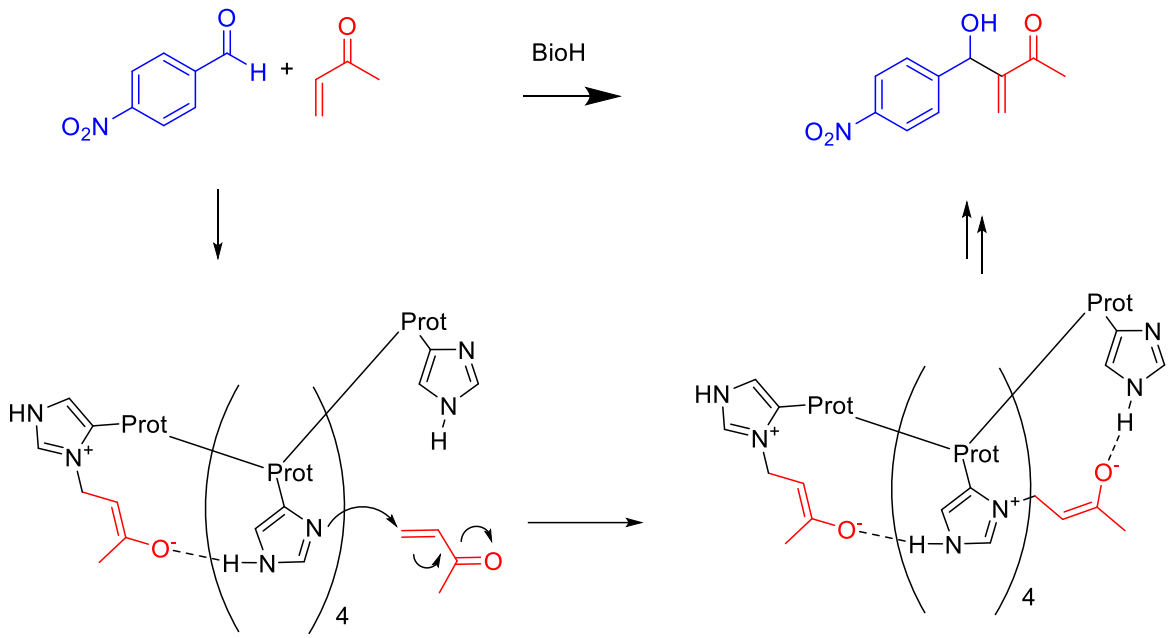


Figure 2.50: Proposed first step of the hexahistidine-tag catalysed MBH reaction. The oxyanion formed in the first step of the MBH reaction could be stabilised by a hydrogen-bond with neighbouring histidine of the tag

To test this hypothesis, variants of EcBioH WT with different numbers of hexahistidine tags (0, 1 and 2) will be produced and assayed. While both variants of EcBioH with one hexahistidine tag (EcBioH-N and EcBioH-C) were already described earlier, the untagged version of EcBioH (EcBioH-x)

has been cloned already. Hence, at this point, only the version of EcBioH with two hexahistidine tags (EcBioH-NC) was to be produced. This was made by cutting the C-terminal hexahistidine tagged EcBioH gene, using Nde1 and Xho1 restriction enzymes, from a pET22a plasmid and ligating it to a pET28a plasmid. EcBioH-NC was purified following the same procedure than the one described in part 2.1.2. The only change observed is that the EcBioH-NC eluted later, at a concentration of 400 mM imidazole, in IMAC, as expected. After SEC, yields of 14 mg/L of culture of EcBioH-NC were obtained. The monomeric enzyme was determined by SDS-PAGE (shown on Figure 2.51) to have 95+% purity.

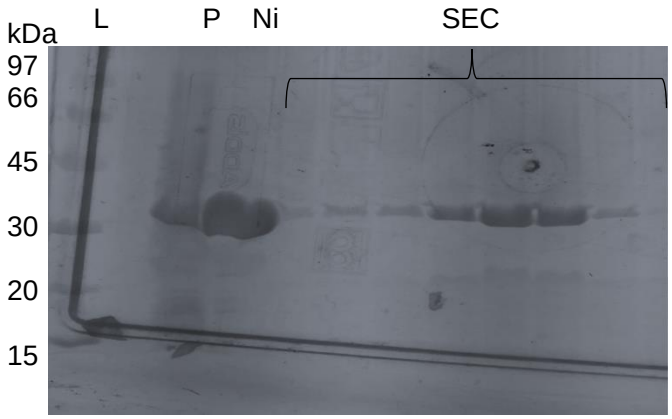


Figure 2.51: SDS-PAGE gel (12%) of EcBioH-C purification steps
L: Low molecular weight ladder; Ni: Sample post-Ni affinity chromatography; SEC: Sample post-size-exclusion chromatography.

EcBioH-x was purified first by anion exchange chromatography, using a 0 to 1 M sodium chloride at pH 7.5 gradient and EcBioH-x eluted at 150 mM NaCl. SDS-PAGE analysis showed that the fractions containing the protein had a purity of 60+%. Notably, two proteins with molecular weights of 50 to 60 kDa were observed. Indeed, as EcBioH-x pI as calculated by the ExPASy-ProtParam tool is 6.5, hence very close to biological pH, it was challenging to obtain a higher purity from the ion exchange step, thus enhancing the need of the purification by SEC. After SEC, some fractions still contained contaminant proteins but other fractions, which were used for further analysis, were >90% pure as showed on the SDS-PAGE in Figure 2.52.

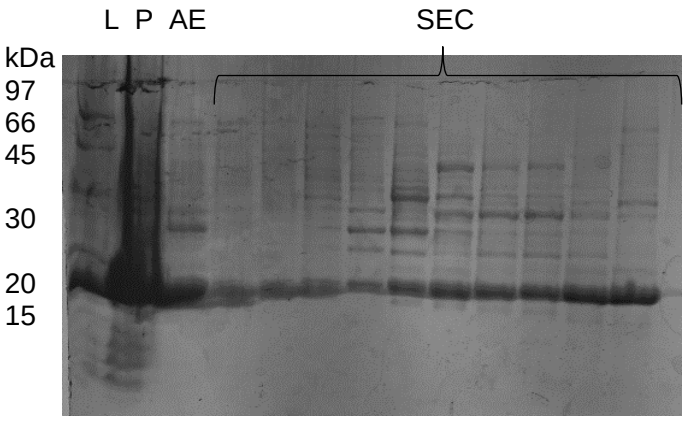


Figure 2.52: SDS-PAGE (12%) gel of EcBioH-x purification steps
L: Low molecular weight ladder; Ni: Sample post-Ni affinity chromatography; SEC: Sample post-size-exclusion chromatography.

Both forms were also characterised by LC-MS(ESI). The spectra and the results are showed in Figure 2.53 and Table 2.14:

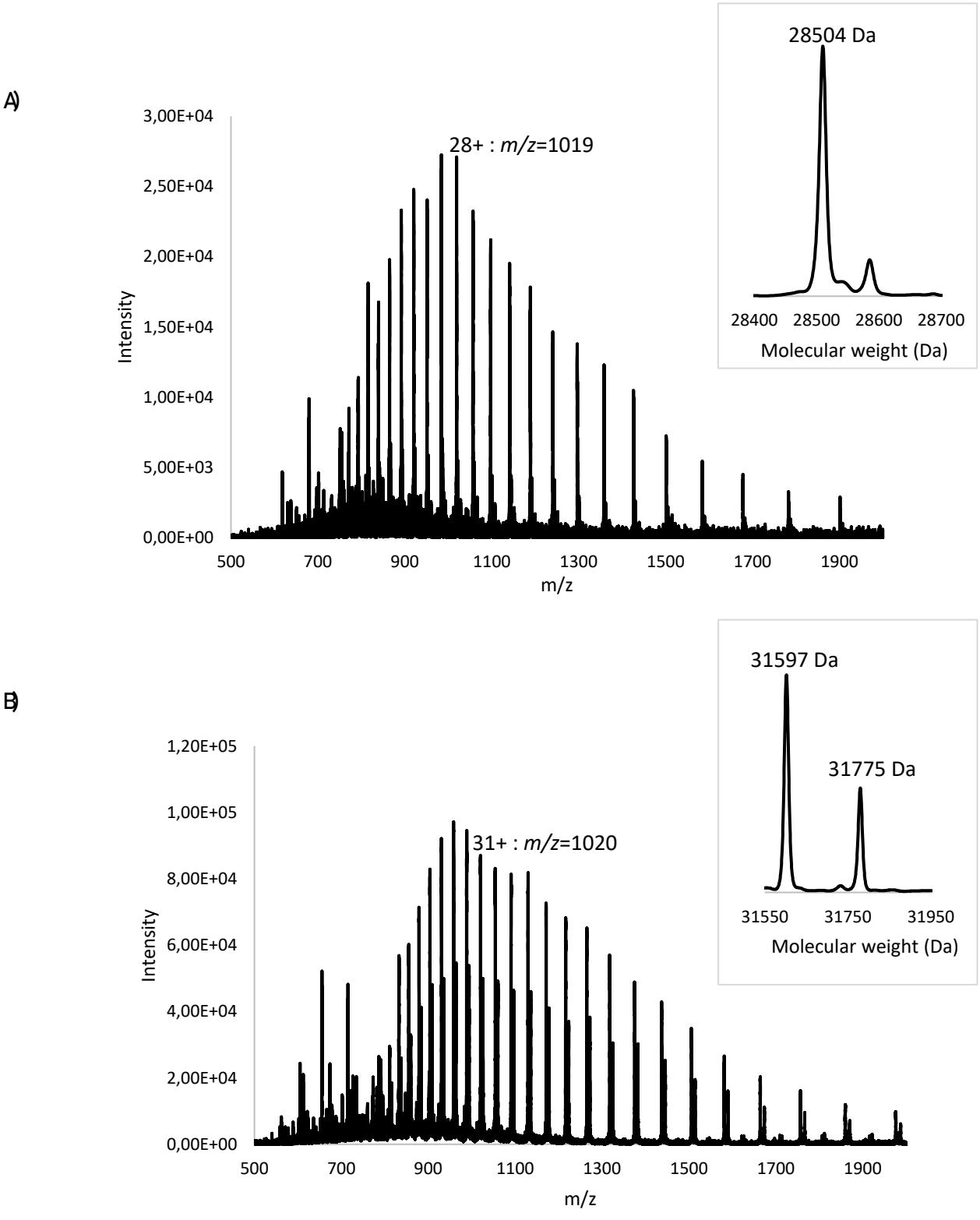


Figure 2.53: ESI-MS spectra of the EcBioH clones expressed:
A) EcBioH-x
B) EcBioH-NC. A second set of peaks corresponding to the glycophosphorylated protein (m + 178 Da) was also observed

Entry name	Number of hexahistidine tags	Theoretical molecular weight* (Da)	Determined molecular weight (Da)
EcBioH-x	0	28505.01	28503.97 ± 2.87
EcBioH-NC	2	31596.32	31597.02 ± 1.42

Table 2.14: MS data of the EcBioH clones expressed.
** Theoretical molecular weights were calculated by the ExPASy-ProtParam tool. The average mass is shown.*

2.8.5 Assaying EcBioH variants with different number of hexahistidine tags

All these variants of EcBioH were assayed by HPLC for the MBH reaction. The results are shown in Figure 2.54.

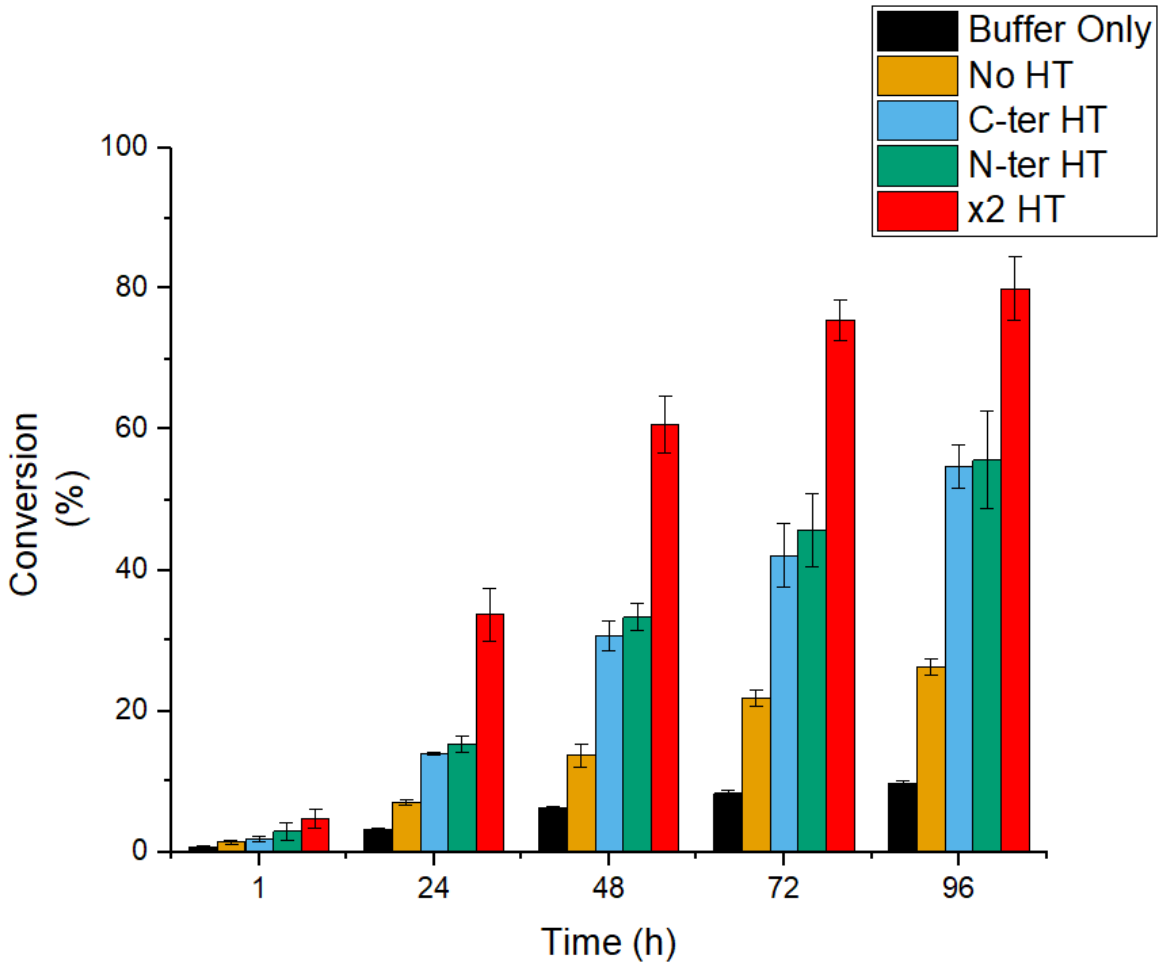


Figure 2.54: Time course of the MBH reactions with variants of EcBioH with different number of hexahistidine tags

It was observed that the presence of a hexahistidine-tag directly impacts the percentage conversion of the MBH reaction. Indeed, while after 24 hours, a conversion of 7% is observed for the EcBioH-x variant, a 2-fold conversion increase was observed for both EcBioH-N and EcBioH-C. Furthermore, a 5-fold conversion increase was noted for EcBioH-NC. In addition, EcBioH-N and EcBioH-C display comparable results. This primary data is shown on Figure 2.55. This suggests an involvement of the hexahistidine tag in the EcBioH-catalysed MBH reaction and gives credit to the mechanism proposed in Figure 2.50.

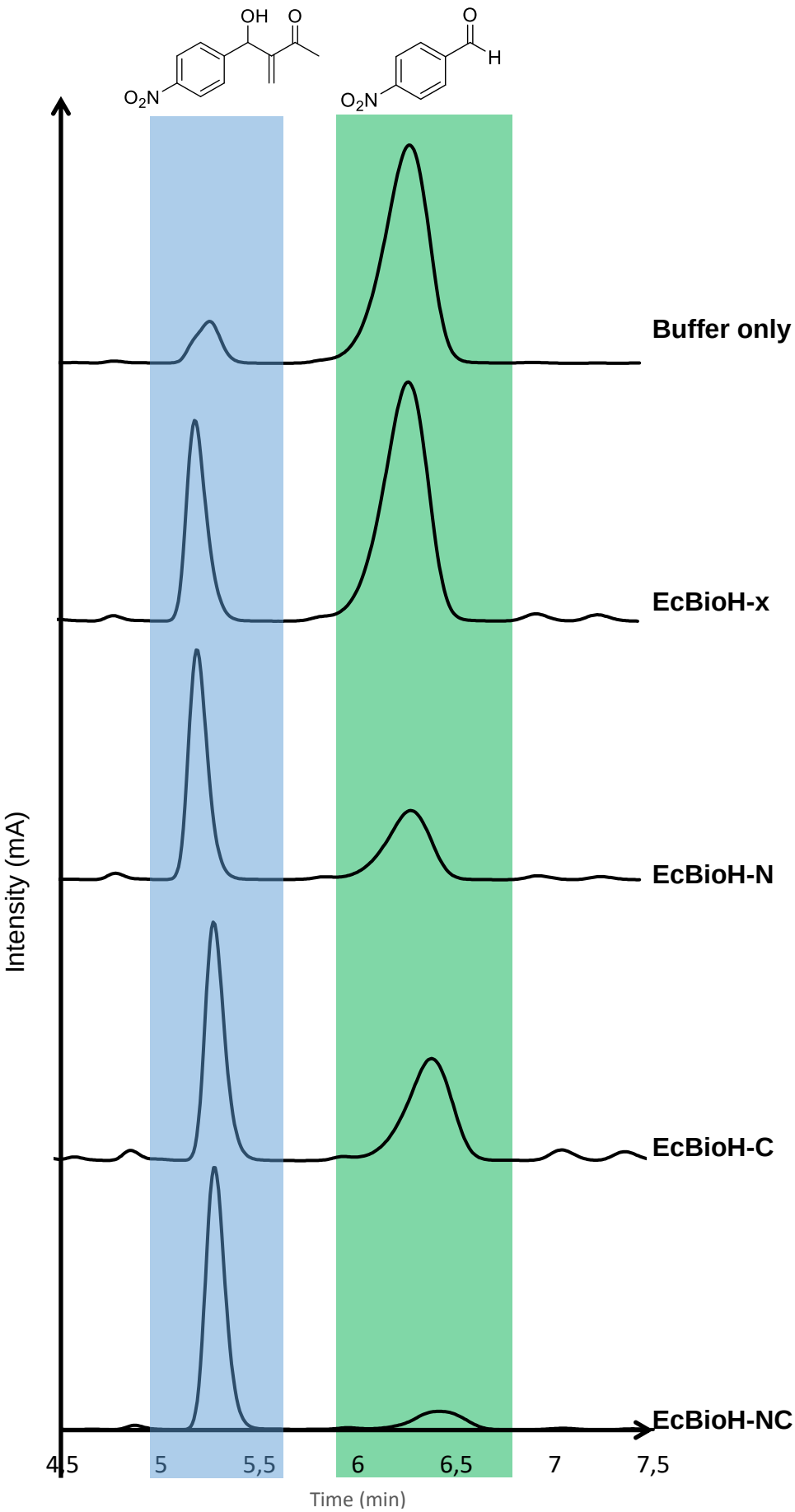


Figure 2.55: HPLC chromatograms of the MBH biocatalysed reaction with various numbers of hexahistidine tags after 96 hours.
Conditions: 4-NBA concentration: 8 mM; MVK: 64 mM; EcBioH-C loading: 130 µM; Temperature: 30C; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5.

2.8.6 Further investigation of the hexahistidine tag in the MBH reaction

With this interesting observation in hand, the next step was to further establish the involvement of the hexahistidine tag through assessing the activity of other proteins both containing a hexahistidine tag or not. As such, various proteins used in the Campopiano lab, which will be briefly introduced next, were assayed for MBH activity.

2.8.6.1 *E. coli* Acyl carrier protein (ACP)

ACP is a well-known peptide carrier protein, involved in various biosyntheses, including fatty acids and polyketides[18]. This protein does not display any enzymatic activity. It is a flexible protein, containing only α-helices and large loops[54]. This protein is however very small, only containing 78 residues for a molecular weight of 8.6kDa[54]. The ACP that was used in this study was from *E. coli.* Its crystal structure was first determined by Roujeinikova and co-workers and is showed in Figure 2.56[54].

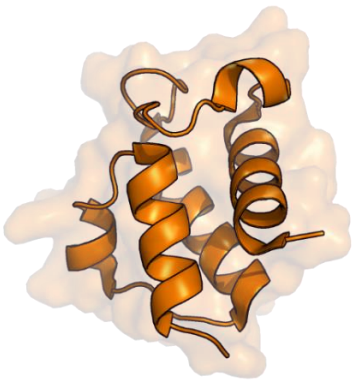


Figure 2.56: Crystal structure of E. coli ACP (PDB ID: 1L0I).

2.8.6.2 *E. coli* 8-amino-7-oxononanoate synthase (BioF)

BioF is the enzyme directly succeeding BioH in the biotin biosynthetic pathway, as displayed in Figure 2.57, and catalyses the pyridoxal 5'-phosphate (PLP)-dependant decarboxylative Claisen-like condensation of pimeloyl-ACP and L-alanine to generate 8-amino-7-oxononanoate[55].

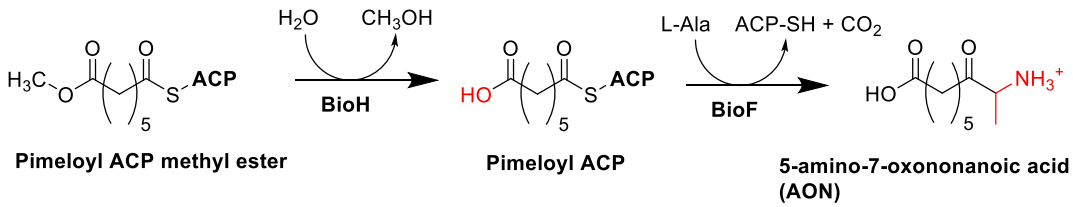


Figure 2.57: BioH-BioF reaction scheme.

With a molecular weight of 41.6 kDa, *E. coli* BioF forms a homo-dimer. Its crystal structure, shown in Figure 2.58, was determined by Alexeev, Campopiano, Baxter and colleagues at the University of Edinburgh in 1998[56]. It shows that, similarly to BioH, BioF contains an α/β-hydrolase fold as five β-sheets are flanked by α-helices in one of the three domains of the enzyme.

Figure 2.58: Crystal structure of E. coli BioF (PDB ID: 1BS0).

2.8.6.3 *E. coli* 3-oxoacyl-ACP-synthase (FabH)

FabH is an enzyme involved in fatty acid biosynthesis. It catalyses the addition of 2 carbons onto malonyl-ACP to form 3-oxobutanoyl-ACP, using acetyl-coenzyme A as a carbon donor, as shown on Figure 2.59[57].

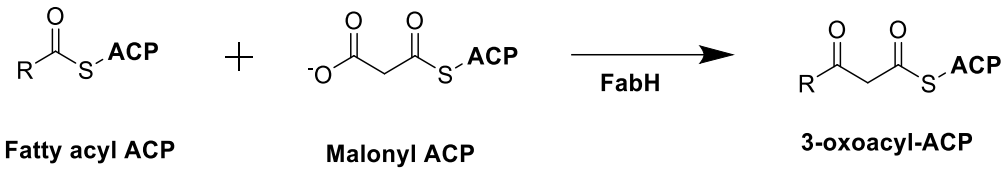


Figure 2.59: FabH reaction scheme.

E. coli FabH is made of 317 residues and has a molecular weight of 33.5 kDa, which compares with EcBioH, 29.5kDa. While the covalent inhibition of FabH by antimicrobial agent 4,5-dichloro-1,2-dithiol-3-one was studied in the Campopiano lab[57], the structure of *E. coli* FabH dimer was determined in 2009 and is showed in Figure 2.60[58].

Figure 2.61: Crystal structure of E. coli FabH (PDB ID: 3IL9).

2.8.6.4 D-phenylglycine aminotransferase (D-PhgAT)

The last enzyme which was investigated was D-PhgAT from *P.stutzeri.* This PLP-dependent enzyme catalyses the amino group transfer from L-glutamic acid to benzoylformate to form D-phenylglycine and ketoglutarate as shown on Figure 2.62[59].

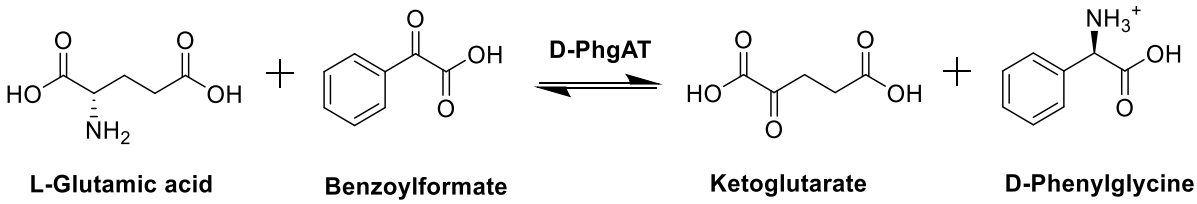


Figure 2.62: D-PhgAT reaction scheme

This enzyme is much larger than EcBioH, with a molecular weight of 49kDa[34]. D-PhgAT was extensively studied in the group and, while the first *apo* crystal structure of this enzyme was deposited in the PDB in 2006[59], the first pyridoxal-5'-phosphate internal aldimine-bound structure was added to the PDB by Serpico, Marles-Wright and Campopiano in 2018 and is showed on Figure 2.63[60].

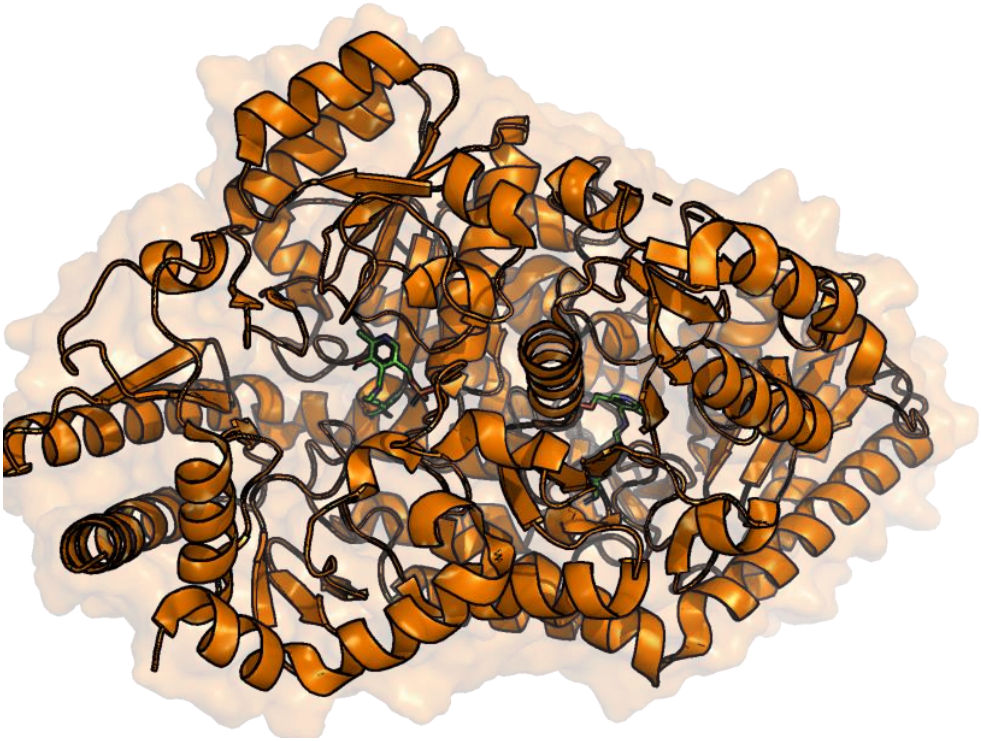


Figure 2.63: Crystal structure of a P.stutzeri D-PhgAT dimer (PLP shown in green) (PDB ID: 6G1F).

2.8.6.5 Assaying these enzymes

Hence, to further investigate the involvement of the hexahistidine tag, the tagged enzymes presented in parts 2.8.6.1-4 were transferred into the reaction buffer by discontinuous diafiltration and were assayed along active EcBioH-C and thermally denatured EcBioH-C (heated at 95°C for one hour). The HPLC-measured conversion is shown in Table 2.15.

Protein	Conversion after 96 hours (%)
EcBioH-C	55 ± 2
Denatured EcBioH-C	53 ± 2
ACP	63 ± 4
BioF	44 ± 7
FabH	57 ± 2
D-PhgAT	42 ± 5

Table 2.15: MBH conversion for various hexahistidine-tagged proteins.
Conditions: 4-NBA concentration: 8 mM; MVK: 64 mM; EcBioH-C loading: 4 mg/mL; Temperature: 30°C; Solvent: 4% MeCN in 20 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.5.

Interestingly, all the hexahistidine tagged proteins show significant and comparable activities for the MBH reaction. Even acyl carrier protein, a peptide carrier protein which does not display any enzymatic capability in the literature, presents a similar activity to EcBioH-C. Besides, denaturation does not affect EcBioH-C's catalytic activity for the MBH reaction as they both have similar activities.

These results further provide strong evidence both that the fold of the enzyme does not have any positive effect on the EcBioH catalytic efficiency and that any hexahistidine-tagged proteins would display significant MBH catalytic activity.

2.9 A general residue catalysis?

It was proven that the hexahistidine tag provides a significant improvement over the untagged protein, as, for EcBioH-x, the conversion after 24 hours was 7% for the untagged protein, 15% for the hexahistidine tagged protein and 34% for the double hexahistidine-tagged protein. However, some activity is still observed for the untagged protein. This activity might be due to unspecific residue catalysis from the protein, including by the histidine residues contained in the sequence. To prove this, EcBioH-x was assayed alongside denatured EcBioH-x and untagged bovine serum albumin (BSA). BSA being a serum albumin protein, it does not have any enzymatic activity. The results of this assay are shown in Table 2.16:

Protein	Conversion after 96 hours (%)
EcBioH-x	26 ± 1
Denatured EcBioH-x	25 ± 3
BSA	34 ± 4

Table 2.16: MBH conversion for various untagged proteins.

Hence, similarly to the previous results for hexahistidine tagged enzymes, denatured EcBioH-x and non-enzymatic proteins show similar activities to folded EcBioH-x. This is a strong evidence that significant activity results from unspecific residue catalysis.

2.10 Conclusion

This study aimed at gaining mechanistic insights into the biocatalysed MBH reaction. After choosing a suitable platform for that study in EcBioH and determining optimised conditions for the reaction, the origin of the EcBioH catalysis was investigated. It was found that part of the activity is caused by unspecific residue catalysis, while a significant increase in activity is observed when a

hexahistidine tag is present. It is postulated that this increase in activity is explained by a synergistic relationship between the oxyanionic intermediates of the MBH reaction and the hexahistidine tag. Indeed, the oxyanionic intermediates could be stabilised by H-bond with a neighbouring histidine, which, in return, will gain electron density and will be able to undergo Micheal addition on the electrophilic alkene more easily as shown on Figure 2.64.

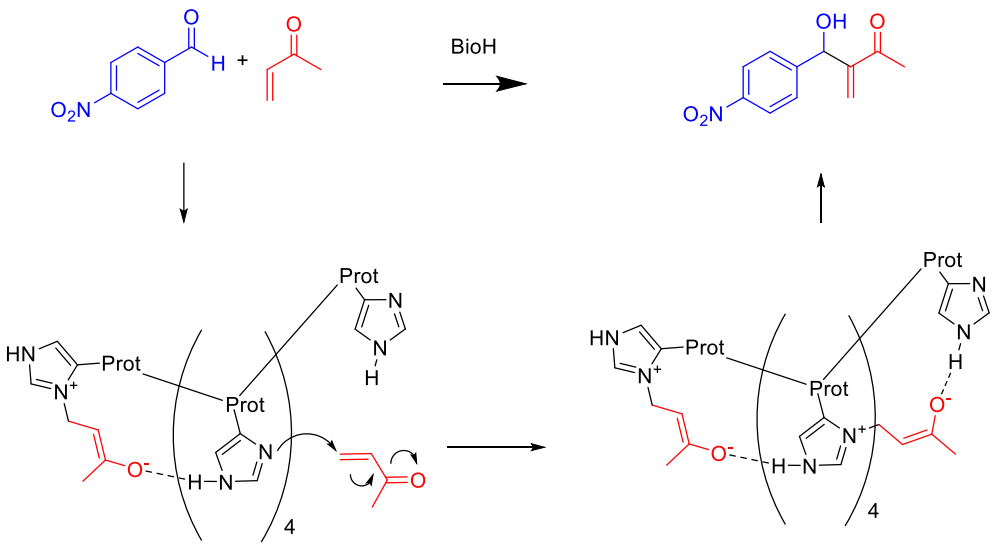


Figure 2.64: Proposed hexahistidine-tag catalysed MBH reaction: the oxyanionic intermediates could be stabilised by H-bond interaction with a neighbouring histidine, which, in return, could gain electron density and allowing it to undergo Micheal addition on the electrophilic alkene.

Further study is needed to provide support to this mechanism, possibly by molecular dynamics simulations, yet this work provides solid data that supports the high impact of the hexahistidine tag on the biocatalysed MBH reaction. It is interesting to note that some of the biocatalysts investigated in the previous literature contain a hexahistidine tag[24] and some do not[26]. In some other reported cases, the presence or not of a hexahistidine tag is not specified[25]. In this project, it was observed that a hexahistidine tag increases the MBH conversion 2-fold after 24 hours. With this in mind, it becomes hard to compare the previously published biocatalysts.

Hence, this work provides strong evidence against some previously hypothesised mechanisms and most notably against the involvement of a catalytic triad or of a cysteine residue in the biocatalysed reaction. This study has also shed light on the significant impact of the hexahistidine tag in MBH biocatalysis and shows that general residue catalysis is observed as any protein, folded or denatured, displays significant MBH catalytic activity.

References

[1] K. Morita, Z. Suzuki *et al.*, 1968, *Bull. Chem. Soc. Jpn.,* 41, p 2815

[2] M. Shi, F. Wang *et al.,* 2011, *The Chemistry of the Morita-Baylis-Hillman reaction,* RSC Publishing, Cambridge, UK

[3] V.K. Aggerwal, S.Y. Fulford and G.C. Lloyd-Jones, 2005, *Angew.Chem.Intl.Ed.,* 44 pp 1706-1708

[4] S.E. Drewes, G.H.P. Roos, 1988, *Tetrahedron*, 44 pp 4653-4670

[5] A. Gilbert, T.W. Heritage, N.S. Isaacs, 1991, *Tetrahedron: Asymmetry,* 2 pp 969-972

[6] L.J. Brzezinski, S. Rafael, J.W. Leahy, 1997, *J.Am.Chem.Soc.*, 119 pp 4317-4318

[7] T. Hayase, T. Shibata *et al.,* 1998, *Chem. Comm.,* pp 1271-1273

[8] Y. Iwabuchi, M. Nakatani *et al.*, 1999, *J.Am.Chem.Soc,* 121 pp 10219-10220

[9] Y. Sohtome, N. Takemura *et al.,* 2008, *Tetrahedron,* 64 pp 9423-9429

[10] N. McDougal, S.E. Schaus, 2003, *J.Am.Chem.Soc.,* 125 pp 12094-12095

[11] M.M. Vasbinder, J.E. Imbriglio, S.J. Miller, 2006, *Tetrahedron,* 62 pp 11450-11459

[12] Q. He, G. Zhan *et al.,* 2016, *Beilstein J.Org.Chem,* 12 pp 309-313

[13] H. Yoshimura, K. Takahashi *et al.*, 2015, *Chem.Comm.,* 51 pp 17004-17007

[14] Z. Duan, Z. Zhang *et al.*, 2013, *RSCAdv.,* 3 pp 10127-10130

[15] C. Wang, L. Wang, *et al.,* 2013, *Phosphorus, Sulfur, Silicon,Relat.Elem.,* 188 pp 1548-1554

[16] M. Reetz, R. Mondière *et al.,* 2007, *Tetrahedron Lett.,* 48 pp 1679-1681

[17] S. Bjelic, L.G Nivon *et al.,* 2013, *ACS Chem.Bio.,* 8, pp 749-757

[18] P.N. Joshi, L.Purushottam *et al.,* 2016, *RSC Adv.,* 6, pp 208-212

[19] L. Jiang, H.W. Yu, 2013, *Biotechnol.Lett,* 36 pp 99-103

[20] L. Jiang, H.W. Yu, 2014, *Chem.Res.Chin.Univ.,* 30 pp 289-292

[21] J.T. Mindrebo, C.N. Nartey *et al.,* 2016, *Curr.Op.Structur.Biol.,* 41 pp 233-246

[22] N.H. Tomczyk, J.E. Nettleship *et al.,* 2002, *FEBS Lett.,* 513 pp 299-304

[23] J.E. Cronan, 2001, *J.Biol.Chem,* 276 pp 37355-37364

[24] S. Lin, J.E. Cronan, 2012, *J.Biol.Chem.*, 287 pp 37010-37020

[25] S. Lin, R.E. Hanson, J.E. Cronan, 2010, *Nature Chem.Bio.,* 6 pp 682-688

[26] W.R. Streit, P. Entcheva, 2002, *Appl.Microbiol.Biotechnol.,* 61 pp 21-31

[27] W. Wei, H. Guan *et al.*, 2019, *iScience*, 19, pp 796-808

[28] Q. Zeng, Q. Yang *et al.,* 2020, *Front.Microbiol.,* 11, 148

[29] W. Salaemae, A. Azhar *et al.,* 2011, *Prot. Cell.,* 2 pp 691-695

[30] R. Sanishvili, A.F. Yakunin *et al.,* 2003, *J.Biol.Chem.,* 278 pp 26039-26045

[31] V. Agarwal, S. Lin, *et al.,* 2012, *Proc.Natl.Acad.Sci.USA,* 109 pp 17406-17411

[32] J.L. Arpigny, K.E. Jager, 1999, *Biochem.J.,* 343 pp 177-183

[33] L. Wang, Y. Chen *et al.*, 2019, *Biochem.Biophys.Res.Commun.,* 520, pp 538-543

[34] J. Shi, X.Cao *et al.*, 2016, *Biochem.,* 55 pp 6705-6717

[35] S.R. Harrowing, J.B. Chaudhuri, 2003, *J.Biochem.Biophys.Met.* 56, pp 177-188

[36] J. Clayden, N. Greeves *et al.*, 2000, *Organic Chemistry, 1st Edition,* Oxford University Press, Oxford, UK

[37] M. Lopez-Iglesias, E. Busto, *et al.,* 2011, *Adv.Synth.&Catal.* 353, pp 2345-2350

[38] H. Liu, J.H. Naismith, 2008, *BMC Biotechnol.,* 8 doi:10.1186/1472-6750-8-91

[39] S.R. Trevino, J.M. Scholtz, C.N. Pace, 2007, *J.Mol.Biol.,* 366, pp 449-460

[40] I. Hunt, *Table of pKa and pI values,* accessed from http://www.chem.ucalgary.ca/courses/351/Carey5th/Ch27/ch27-1-4-2.html on 18/09/19

[41] A. Domsalla, M.F. Melzig *et al.* 2008*, Planta Med.,* 74, pp 699-711

[42] G. Ferrer-Sueta, B. Manta, 2011, *Chem.Res.Toxicol.*, 24, pp 434-450

[43] S.M. Marino, V.N. Gladyshev, 2010, *J.Mol.Biol.,* 404, pp 902-916

[44] A. Narayanan, L.H. Jones, 2015, *Chem.Sci.*, 6, pp 2650-2659

[45] K. Chibale, 2002, *Arch.Org.Chem,* 9, pp 93-98

[46] X. Zhang, H. Du *et al.,* 2006, *J.Org.Chem.*, 71, pp 2862-2869

[47] L.R. Domingo, J. Andres, 2003, *J.Org.Chem.,* 68, pp 8662-8668

[48] P.M. Pihko, 2004, *Angew.Chem.Intl.Ed.,* 43 pp 2062-2064

[49] S. Luo, B. Zhang *et al.,* 2002, *Tetrahedron Lett.,* 43, pp 7369-7371

[50] M. Baidya, F. Brotzel *et al.,* 2010, *Org.Biomol.Chem.,* 8, pp 1929-1935

[51] S. Luo, P.G. Wang, *et al.,* 2004, *J.Org.Chem.,* 69, pp 555-558

[52] M.P. Pereira, R. de Souza Martins *et al.*, 2018, *RSC Adv.,* 8, pp 23903-23913

[53] L. Raich, H. Santos *et al.*, 2018, *ACS Catal.,* 8, pp 1703-1714

[54] A. Roujeinikova, C. Baldock *et al.,* 2002, *Structure,* 10 pp 825-835

[55] M. Manandhar, J.E. Cronan, 2017, *App.Environ.Microbiol.,* 84, e02084-17

[56] D. Alexeev, M. Alexeeva *et al.*, 1998, *J.Mol.Biol.,* 284, pp 401-419

[57] A.G. Ekstrom, V. Kelly *et al.,* 2017, *Org.Biomol.Chem.,* 30, pp 6310-6313

[58] K.S. Gajiwala, S. Margosiak *et al.*, 2009, *FEBS Lett.,* 583, pp 2939-2946

[59] C.J.W. Walton, F. Thiebaut *et al.,* 2018, *Biochem.,* 57, pp 5437-5446

[60] A. Serpico, J. Marles-Wright, D.J. Campopiano, 2019, submitted

Chapter 3 Overcoming evolution to expand the synthetic scope of a nitrile synthetase

3.1 Introduction

3.1.1 Nitrile synthetase, an interesting family of enzymes

The nitrile synthetase (NS) family of enzymes is a small group of enzymes catalysing the conversion of 7-carboxydeazaguanine (CDG) into 7-cyanodeazaguanine ($PreQ_0$) in the biosynthetic pathways of queuosine and toyocamycin[1]. The family is magnesium adenosine triphosphate (MgATP)-dependent and was shown to proceed *via* the formation of the corresponding amide intermediate, 7-amidodeazaguanine (ADG)[1]. Interestingly, these enzymes bind a Zn^{2+} cation; however, this zinc is thought to be "structural" and not involved in activity[1]. The scheme of the reaction catalysed by NS is shown in Figure 3.1.

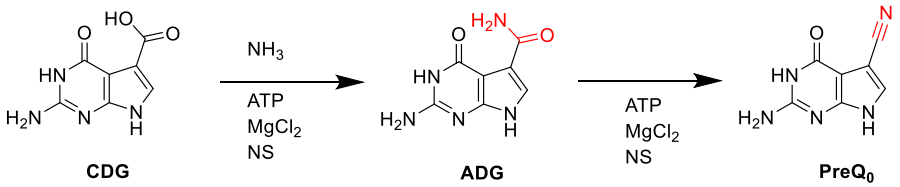


Figure 3.1: NS reaction scheme. CDG is first transformed into ADG, using one equivalent of MgATP and then into $PreQ_0$ using a second equivalent of MgATP.

The reaction has been proposed to proceed *via* the formation of two adenylate intermediates to activate both the acid and then the amide to form the final product, as shown in Figure 3.2[1].

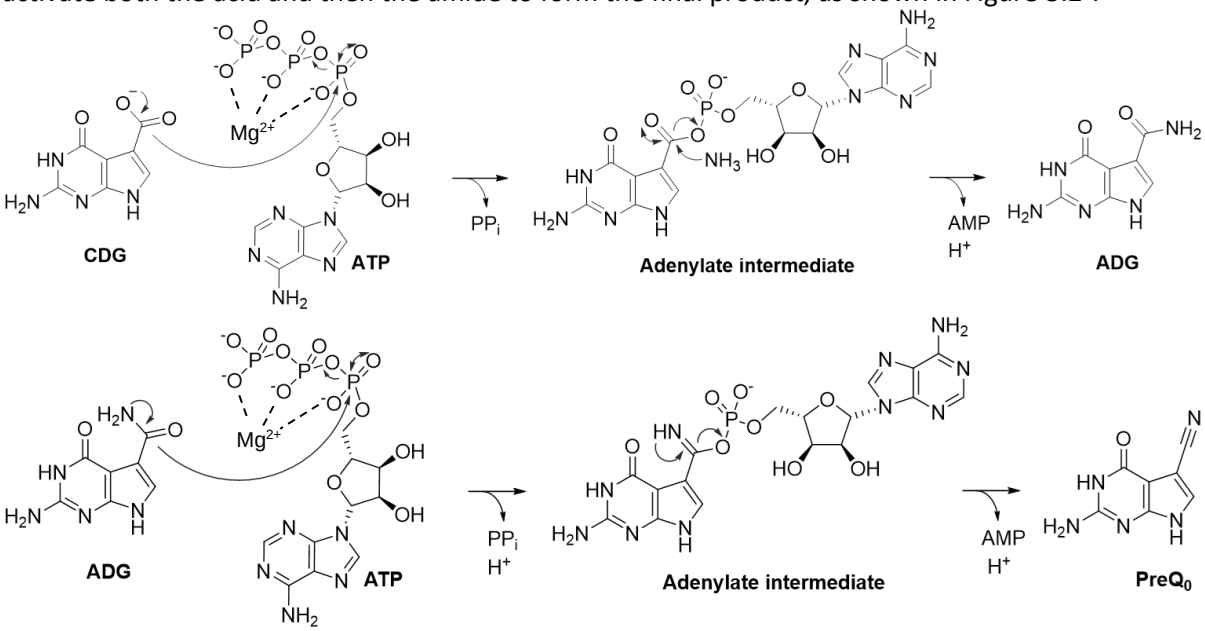


Figure 3.2: NS proposed reaction mechanism. Both transformations are thought to involve the formation of an adenylate intermediate, which will activate CDG in the first step and ADG in the second step.

Thus, the conversion of acid to nitrile uses 2 equivalents of ATP. In the first step of the reaction, the ATP α-phosphate is attacked by CDG carboxyl. This will form a reactive adenylate intermediate that will be broken down by the amine. This is the typical mechanism of the class of amide synthetase, the amide bond synthetases (ABS) the NS belong to. What makes the NS enzymes much less common is the second step: the transformation of the primary amide into the corresponding nitrile. This transformation is thought to proceed *via* a similar mechanism. While no mechanistic insights are currently available, it could be postulated that the amide attacks in a similar way the ATP α-phosphate. Amides being way less reactive than carboxylic acids, it is possible that one or both hydrogens bound to the amide nitrogen engage in H-bonds with enzyme residues and/or with the carboxyl that is part of ADG deazaguanine ring. These H-bond interactions might be able to increase enough the electron density of the amide so that it is capable of attacking the ATP α-phosphate to initiate the reaction. Attempts at separating the two parts of the reaction or at isolating the amide intermediate have not been successful so far. The structure of the active site of the enzyme is also unknown. In fact, it is unknown whether these enzymes have one active site where both the amide coupling and the dehydration of the amide take place or two active sites, each specific for one part of the reaction. Winkler and co-workers postulated that an aromatic residue such as a tyrosine may be involved in binding CDG through π-stacking interactions while an aspartic acid residue could direct the substrate through binding the primary and secondary amines of the substrate pyrimidine ring[1]. However, as no experimental or computational validation was reported, more data is required to reach any conclusion with regards to this hypothesis. As such, little is known about this class of enzymes, yet they can have interesting biocatalytic applications for the syntheses of commercially-relevant amide- or nitrile-containing compounds.

3.1.2 NS in natural products biosynthesis

NS enzymes are involved in the biosyntheses of queuosine and toyocamycin, shown on Figure 3.3.

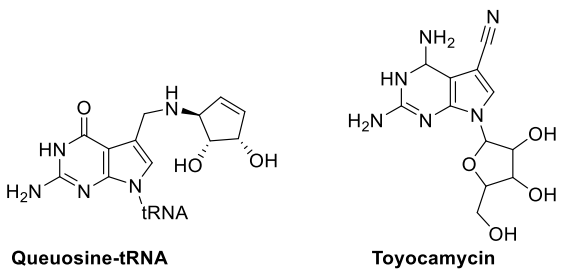


Figure 3.3: Structures of queuosine and toyocamycin.

Queuosine is a modified nucleoside found in tRNAs coding for asparagine, aspartic acid, histidine and tyrosine in bacteria, such as *E. coli* or *B. subtilis* and eukaryotes[2]. Queuosine plays a role in maintaining translational efficacy and improving codon recognition[3]. Low levels of queuosine-tRNAs were found to be associated with cell proliferation and cancer[3]. In contrast, toyocamycin is an antibiotic and antifungal agent produced by many *Streptomyces* bacteria, such as *S. rimosus*[4], *S. toyocaensis*[5] or *S. diastatochromogenes*[6]. Toyocamycin went through phase I trials for its antibiotic properties but did not display significant clinical response and was also found to be a candidate for multiple myeloma, as this compound was found to inhibit endoplasmic reticulum stress-induced mRNA splicing. However, no clinical investigation have been undertaken to develop this indication[7].

Both queuosine and toyocamycin possess a similar pyrrolopyrimidine core with different substituents. Both are biosynthesized from guanosine triphosphate (GTP), and the first four steps of their biosynthetic pathways, to $PreQ_{0,}$ are analogous, and then diverge. The first step of this pathway is the conversion of guanosine triphosphate (GTP) into dihydroneopterin triphosphate (H_2NTP) catalysed by GTP cyclohydrolase I (GCH I)[8]. This enzyme, a member of the cyclohydrolase family, catalyses the hydrolytic ring-opening of the imidazole part of GTP and the elimination of one carbon as formic acid[9]. The following enzymes in these pathways, QueD/ToyB, are lyases that generate 6-carboxy-5,6,7,8-tetrahydropterin (CPH_4)[10]. 7-carboxy-7-deazaguanine synthase (QueE/ToyC) then catalyses a key S-adenosyl methionine (SAM)- and magnesium-dependent ring contraction step, which is initiated by the reductive cleavage of SAM, generating a reactive 5'-deoxyadenosyl radical. This step generates CDG, which is then converted into the nitrile $PreQ_0$ by QueC/ToyM. Thereafter, the two pathways diverge, and two enzymes, specific to the queousine biosynthetic pathway will afford queuosine while four additional enzymes are required to generate toyocamycin, as shown in Figure 3.4.

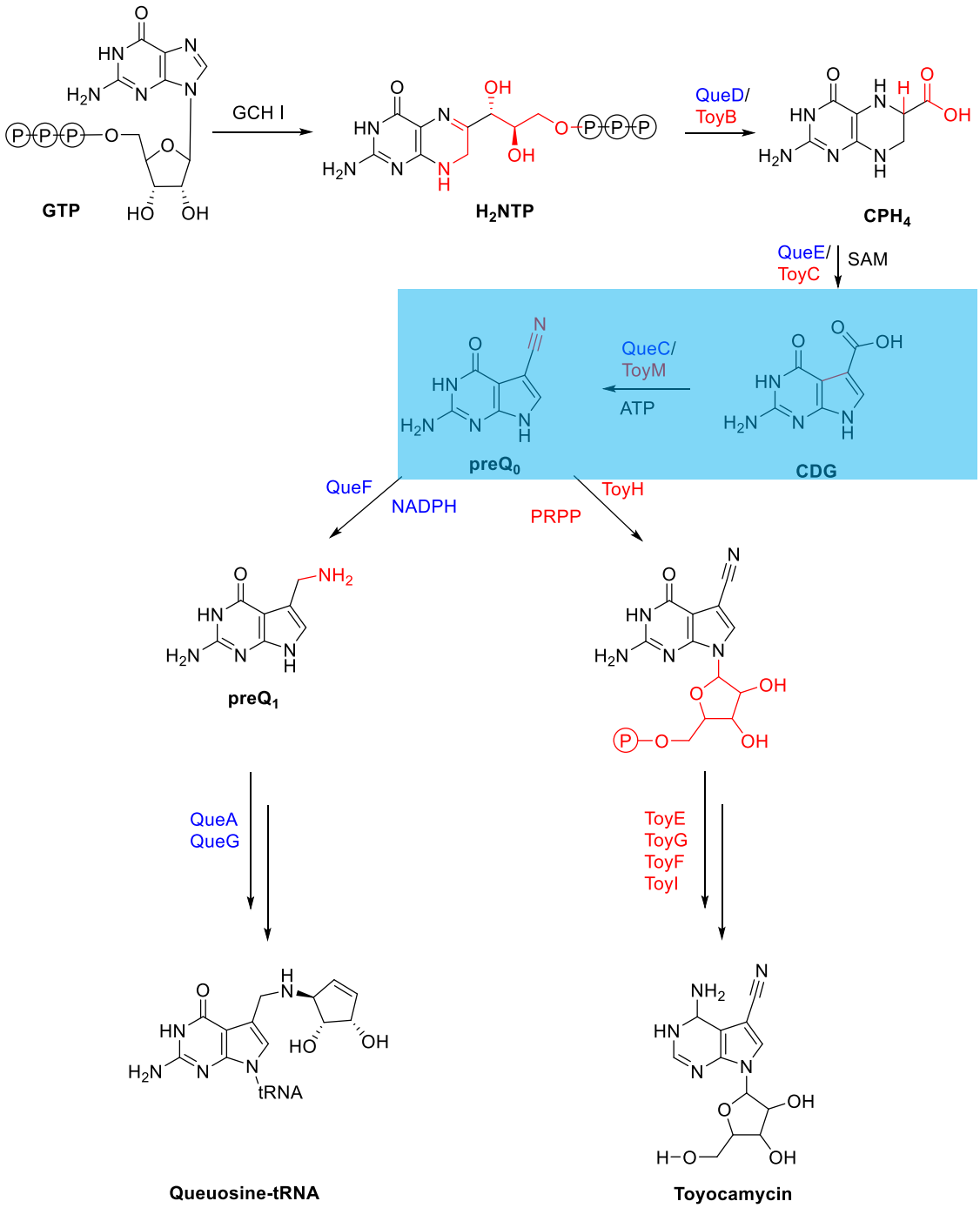


Figure 3.4: Queuosine biosynthetic pathway. Enzymes involved in queuosine biosynthesis are shown in blue and enzymes participating in toyocamycin biosynthesis are shown in red. Both queuosine and toyocamycin biosyntheses use GTP as a starting material. In addition, the first four steps of both pathways are analogous and are catalysed by GTP cyclohydrolase I (GCH I); a lyase (QueD/ToyB); a 7-carboxy-7-deazaguanine synthase (QueE/ToyC) and a NS (QueC/ToyM). Then both biosynthetic pathways diverge.

Interestingly toyocamycin can go through another transformation in some *Streptomyces* species including *S. rimosus* and be transformed in sangivamycin, a potent protein kinase C inhibitor with K_i values in the low micromolar region[11]. Nitrile hydratase ToyJ was found to catalyse the hydration of toyocamycin nitrile into a primary amide to form sangivamycin following the scheme shown in Figure 3.5.

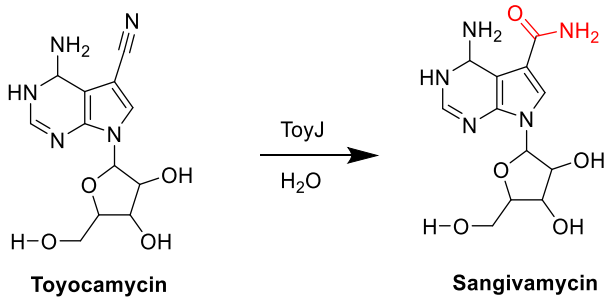


Figure 3.5: ToyJ reaction scheme. This enzyme catalyses a nitrile hydration reaction.

ToyJ is a cobalt-dependent metalloenzyme. It was proposed that ToyJ displays a typical nitrile hydratase mechanism, where the substrate and cysteine sulfenic acid both coordinate to the cobalt for activity following the mechanism shown in Figure 3.6[12].

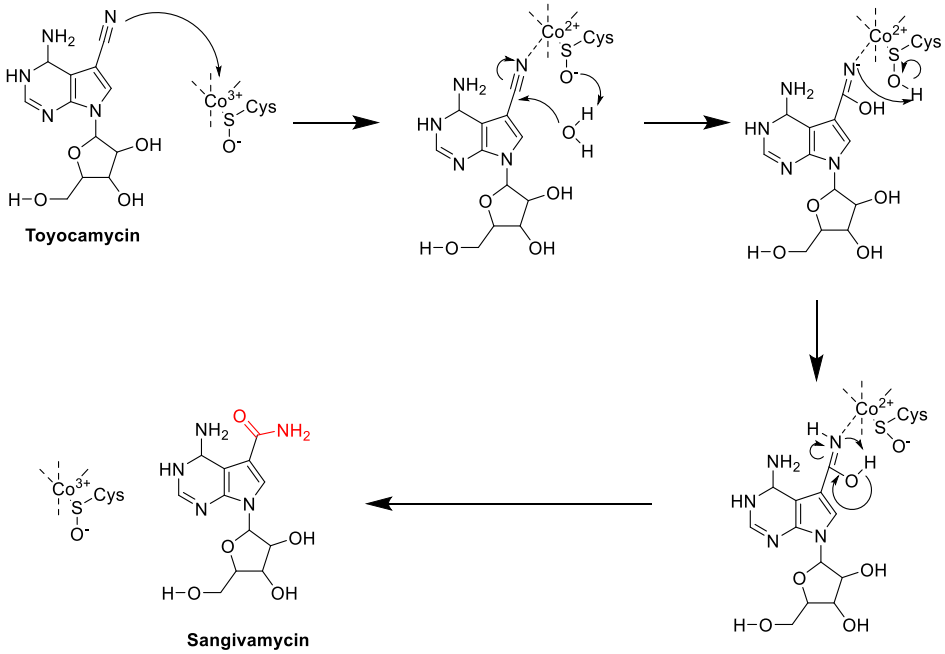


Figure 3.6: ToyJ reaction mechanism. ToyJ is a cobalt-dependent metalloenzyme. Both the substate and a reactive cysteine sulfenic acid coordinate to the catalytic cobalt. The sulfenic acid will be protonated by a water molecule, allowing the hydroxyl to undertake nucleophilic attack on the nitrile. Proton transfer and regeneration of the catalyst lead to the release of sangivamycin

It is interesting to note that, similarly to ToyM, ToyJ makes a primary amide. However, in ToyM case, the primary amide is only an intermediate for the synthesis of the nitrile ($PreQ_0$) and cannot be isolated. Hence, Nelp and Bandarian were able to report the enzyme-catalysed synthesis of ADG through converting CDG into $PreQ_0$ with ToyM and then using ToyJ to hydrate the nitrile to generate ADG, as shown in Figure 3.8[13].

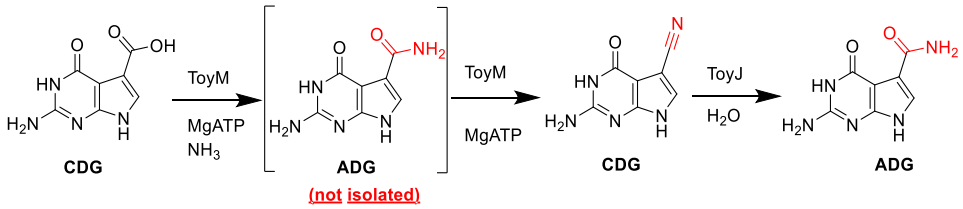


Figure 3.8: ADG biocatalytic route by Nelp and Bandarian[13]: ToyM was first used to convert CDG into $PreQ_0$. While the primary amide was formed in this reaction, the team was not able to isolate it. The nitrile was then converted into the primary amide, using nitrile hydratase ToyJ.

3.1.3 Overview of QueC and ToyM

As previously mentioned, neither QueC nor ToyM mechanism and location of the active site(s) are known. McCarty and co-workers were the first to assay QueC and could observe the production of $PreQ_0$[14]. Winkler and colleagues then found that QueC was extremely specific for its natural substrate, CDG[1]. They have screened a panel of benzoic/indolic acid derivatives for activity with *B. subtilis, E. coli, Archaeoglobus sulficallidus, Pectobacterium atrosepticum* and *Geobacillus kaustophilus* QueC and found that CDG is the only substrate for which any of these enzymes was active[1]. The range of compounds assayed is shown in Figure 3.9:

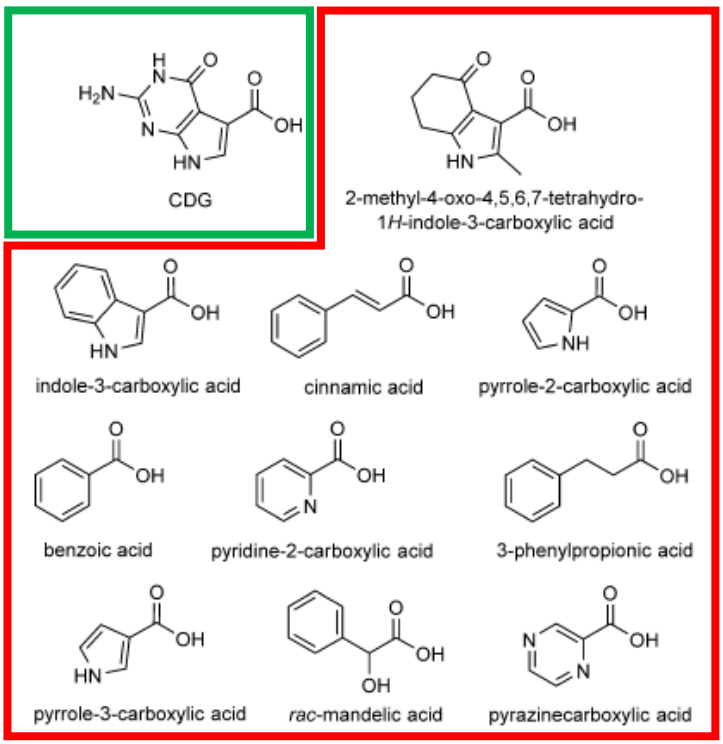


Figure 3.9: QueC substrate screen; Green: Enzyme activity observed, Red: No enzyme activity. Adapted from Winkler et al. [1].

This team was also the first to gain some mechanistic insights on *B. subtilis* QueC (BsQueC) by monitoring the reaction with ^{31}P NMR. Over the time-course, they were able to observe a spectral change that could be consistent with the conversion of MgATP into MgAMP and pyrophosphate[1]. This suggested the formation of an adenylate intermediate.

A sequence alignment of some NS, including those studied by Winkler and colleagues, was also carried out and is shown in Figure 3.10.

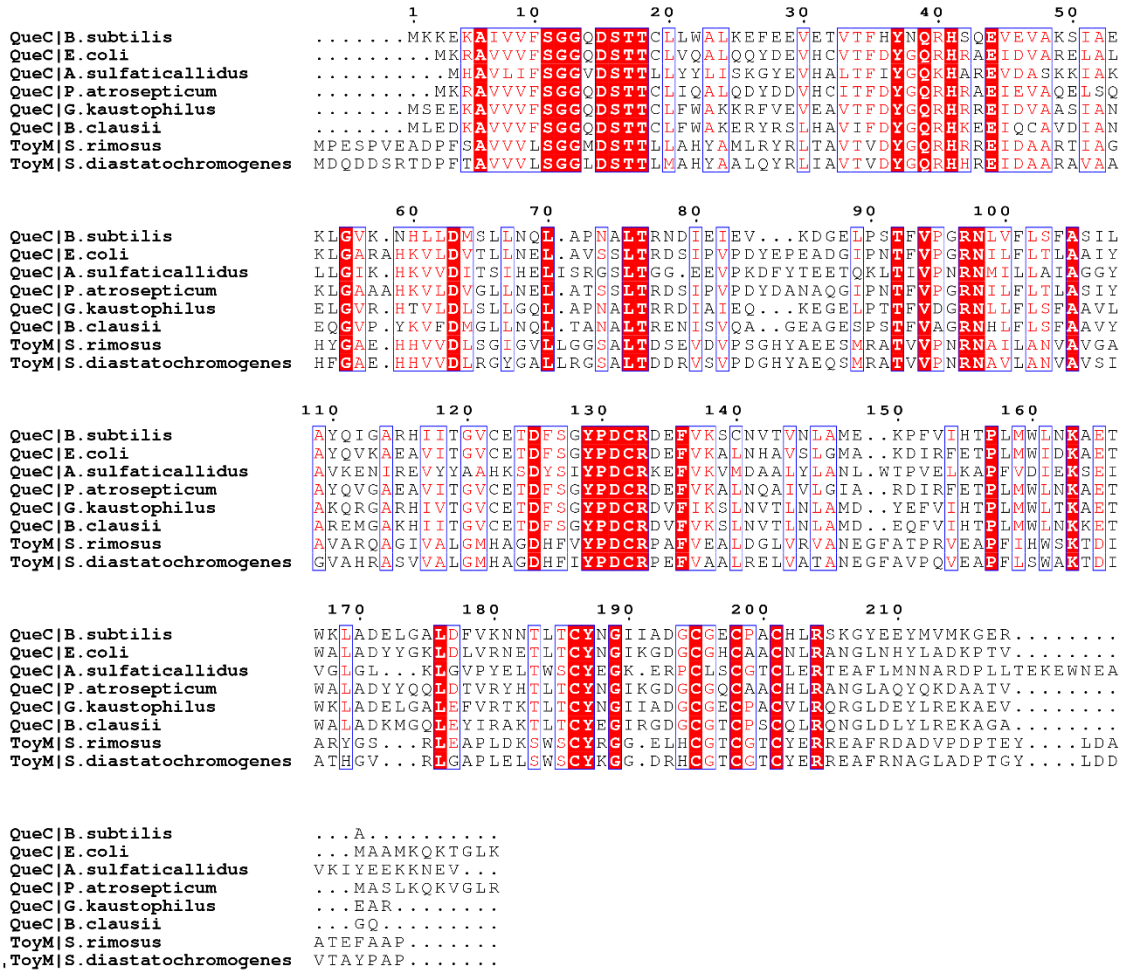


Figure 3.10: Sequence alignment the NSs from B. subtilis (BsQueC) (Uniprot code: O31675), E. coli (Uniprot code: E2QH39), Archeoglobi sulfaticallidus (N0BDM1), Pectobactum atrospecticum (Q6D820), Geobacillus kaustrophilus (Uniprot code:GK1520), B. clausii (Q5WG41), S. rimosus (SrToyM) (Uniprot code: B6CWJ6) and S. diastatochromogenes (A0A1L7B5M4).

Hence, as shown in Figure 3.10, NSs have relatively low sequence identity (20 to 35% to BsQueC). However, this sequence alignment shows that two distinctive features of NS are retained along all species: the proposed Mg^{2+} binding site which corresponds to the well-conserved regions in residues 10-20 and the Zn^{2+} binding site, corresponding to the four conserved cysteines in residues 186-201. Both the known crystal structure, described below, and this sequence alignment will be useful for the protein engineering that will be undergone in this study.

An important milestone in this area was the BsQueC crystal structure, solved in 2008 at a resolution of 2.9 Å, before the protein was first assayed (PDB ID: 3BL5)[15]. *Pectobactum atrospecticum* QueC was also deposited in the PBD in 2007 (PDB ID: 2PG3), without an associated paper, at a resolution of 2.4 Å. While BsQueC was found to form a tetramer in solution, the structure of a BsQueC monomer is shown in Figure 3.11. This structure was determined with both magnesium, phosphate and zinc bound but without CDG or ATP.

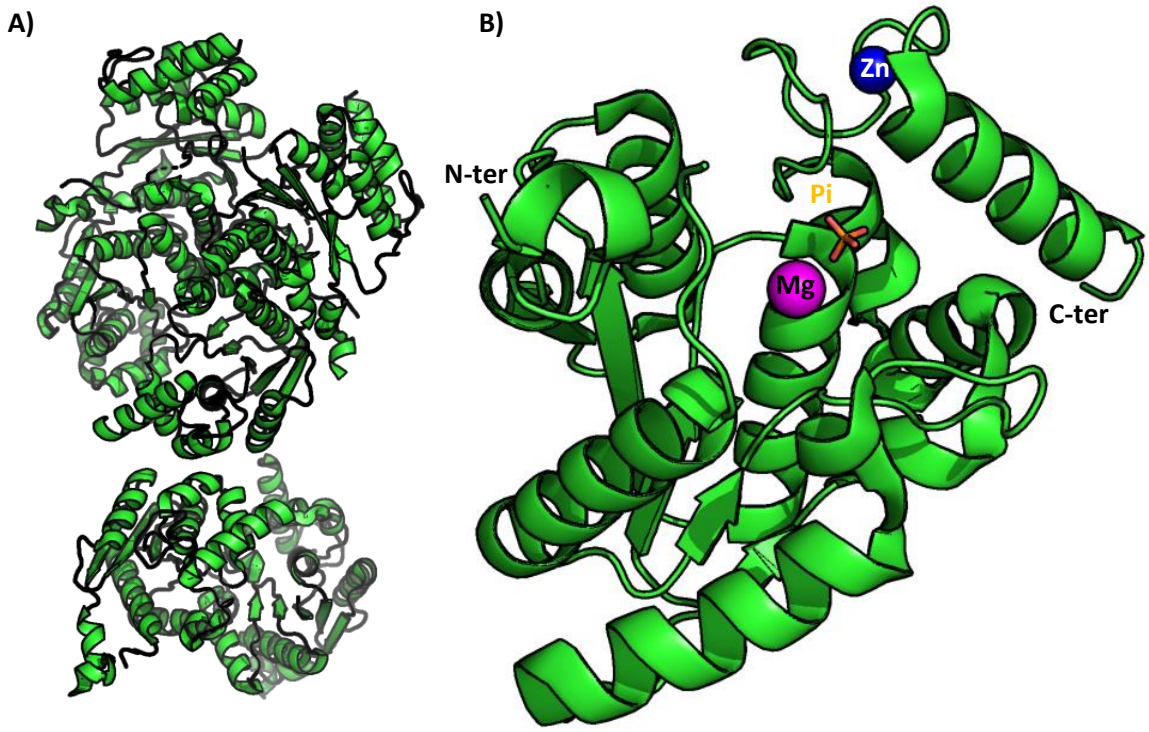


Figure 3.11: A) Structure of the BsQueC tetramer (PDB ID: 3BL5[15]).
B) Structure of a BsQueC monomer Zn^{2+} is shown as a blue sphere and Mg^{2+} is shown as a pink sphere. The phosphate ion is shown in sticks.

Hence, two different parts can be observed in this structure:

- First an N-terminal Rossmann fold composed of five β-sheets flanked by six α-helices. This fold also contains a phosphate-binding loop which is in close proximity to the Mg^{2+} ion and which will help direct the ATP for enzymatic activity. The Rossmann fold is a motif often observed in nucleotide-binding proteins such as BsQueC, as it binds MgATP.
- A C-terminal zinc-binding motif composed of three additional helices which form a Zn-S cluster with four cysteine residues found on one of the helices and on the flexible connecting loops. This Zn-S cluster is shown in close detail in Figure 3.12. However, as mentioned earlier, this zinc is not believed to play a catalytic role in the reaction.

The structure of BsQueC showing the two parts of the structure is shown in Figure 3.13.

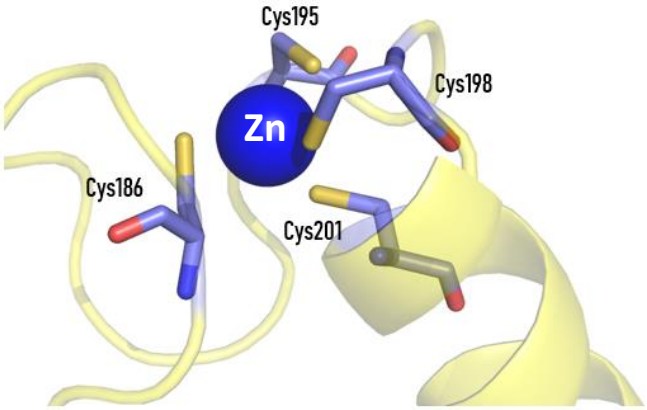


Figure 3.12: Structure of the Zn-Cys complex located on the BsQueC Zn binding site.

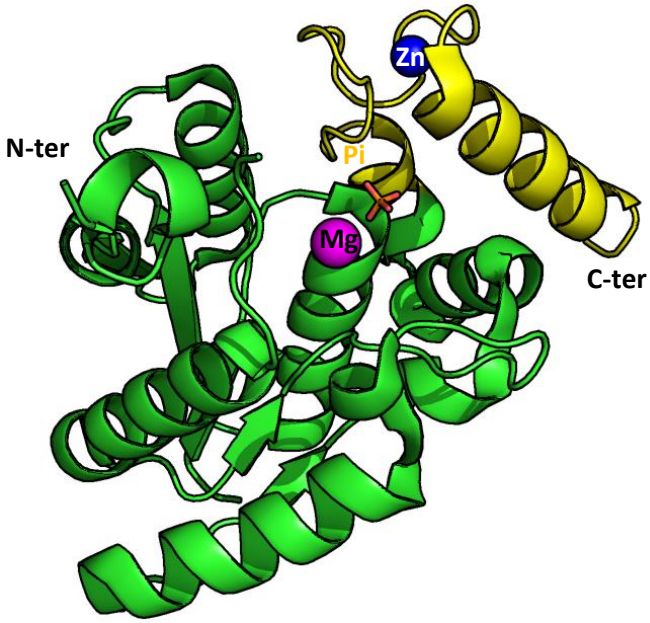


Figure 3.13: BsQueC structure displaying the Rossmann fold in green and the zinc-binding site in yellow. The phosphate ion is shown in sticks.

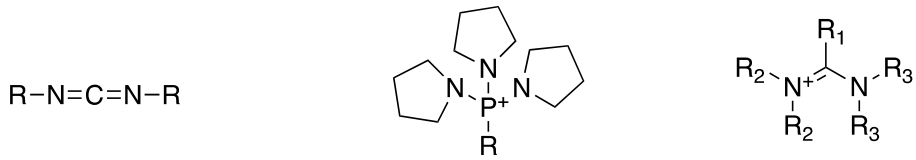
3.1.4 Chemical approaches to amide synthesis from carboxylic acid

As presented earlier, the NS enzymes are able to generate an amide from a carboxylic acid and also to convert that amide into the corresponding nitrile. The generation of an amide from carboxylic acid is the most important reaction in medicinal chemistry by many metrics: in an analysis of 125 manuscripts published in the *Journal of Medicinal Chemistry* in 2014, Brown and Boström have found that 30% of these manuscripts describe at least one acid/amine coupling, making it the most common reaction by far, before Suzuki-Miyaura couplings (20%), aromatic nucleophilic substitutions (14%) or Boc protections/deprotections (9%)[16]. In addition, it is estimated that 25% of all commercial drugs

contain at least one amide group[16]. While the simple coupling of a carboxylic acid and an amine is possible on paper as this reaction is, in general, thermodynamically favorable, a high activation energy is required, hence requiring high temperatures[17]. In addition, the amine will deprotonate the carboxylic acid, hence reducing its reactivity. As such, a great amount of research is dedicated to finding synthetic strategies and reagents to efficiently produce this bond[17]. More specifically, it is about increasing the reactivity of the carboxylic acid through generating a better leaving group than the hydroxyl group on the carboxylic acid. Two of the most common strategies, the use of coupling reagents and the formation of the acyl chloride will be discussed.

3.1.4.1 Coupling reagents

While a large amount of research was dedicated in the development of efficient amide coupling reagents, not a single one of them will be efficient on all substrates. Hence, the choice of coupling reagent can sometimes be a challenge by itself. The role of these reagents is to activate the carboxylic acid and to generate a reactive intermediate with a leaving group good enough for the nucleophilic amine to attack. There are three main types of "go-to" commercial coupling reagents: the carbodiimides, the phosphonium reagents and the aminium/uronium reagents. General structures and examples for these three types of reagents are shown in Figure 3.14.



Carbodiimide	**Phosphonium reagents**	**Aminium/uronium reagents**
Dicyclohexylcarbodiimide (DCC)	**Benzotriazol-1-yloxy-tripyrrolidino-phos- phonium hexafluorophosphate (PyBOP)**	**1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU)**
Diisopropylcarbodiimide (DIC)	Ethyl cyano(hydroxyimino)acetato-O2)- tri-(1-pyrrolidinyl)-phosphonium hexafluo- rophosphate (PyOxim)	N-[(5-Chloro-1H-benzotriazol-1-yl)-dimethylamino-morpholino]-uronium hexafluorophosphate N-oxide (HDMC)
N-(3-Dimethylaminopropyl)-N'-ethylcar- bodiimide · HCl (EDAC)	7-Aza-benzotriazol-1-yloxy-tripyrrolidino- phosphonium hexafluorophosphate (PyAOP)	(1-[1-(Cyano-2-ethoxy-2-oxoethylidene- aminooxy)-dimethylamino-morpholino]- uronium hexafluorophosphate) (COMU)

Figure 3.14: General structure and examples of amide coupling reagents.

The reaction mechanism for the amide coupling, using 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) is shown on Figure 3.15.

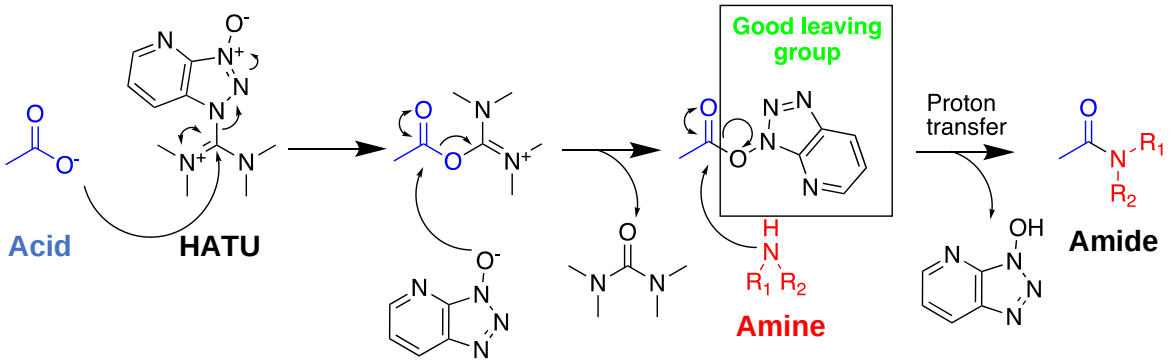


Figure 3.15: Amide coupling mechanism upon using a coupling reagent such as HATU.

Hence, as shown in Figure 3.15, the carboxylic acid derivative first attacks the coupling reagent, to form an unstable *O*-acyl(tetramethyl)isouronium salt, liberating a very nucleophilic anionic triazolopyrimidium N-oxide. This anion will be much more nucleophilic than the amine and, therefore, it will attack the isoronium salt preferentially, forming a triazolopyrimidium ester and liberating a stoichiometric amount of tetramethylurea. Then the amine will finally be able to attack the triazolopyrimidium ester to form the amide and, there again, release a stoichiometric amount of hydroxytriazolopyrimidium.

While HATU is one example of a coupling reagent, all coupling reagents work following relatively similar mechanisms, and all share the same downside - their atom inefficiency. A stoichiometric amount of these coupling reagents is required for the reaction to proceed. In addition, stoichiometric amounts of organic by-products are formed, which may complicate downstream purification. Developing amide coupling reagent recycling systems for liquid- or solid-phase peptide synthesis, to lower the amount of coupling reagent needed, is a growing area of research[18]. However, no widely used viable systems are currently available, and this represents a critical unmet need according to the Americal Chemical Society (ACS) Green Chemistry Institute Pharmaceutical Roundtable held in 2016[19].

3.1.4.2 Making amides via the acyl chloride

Acyl chloride is a very reactive functionality and can be used to produce a wide range of functionalities as shown on Figure 3.16. Acyl chlorides by reacting the corresponding carboxylic acid with one equivalent of thionyl chloride following the mechanism shown in Figure 3.17. Other chlorinating reagents such as phosphorus pentachloride or oxalyl chloride + catalytic amount of dimethylformamide are also commonly used.

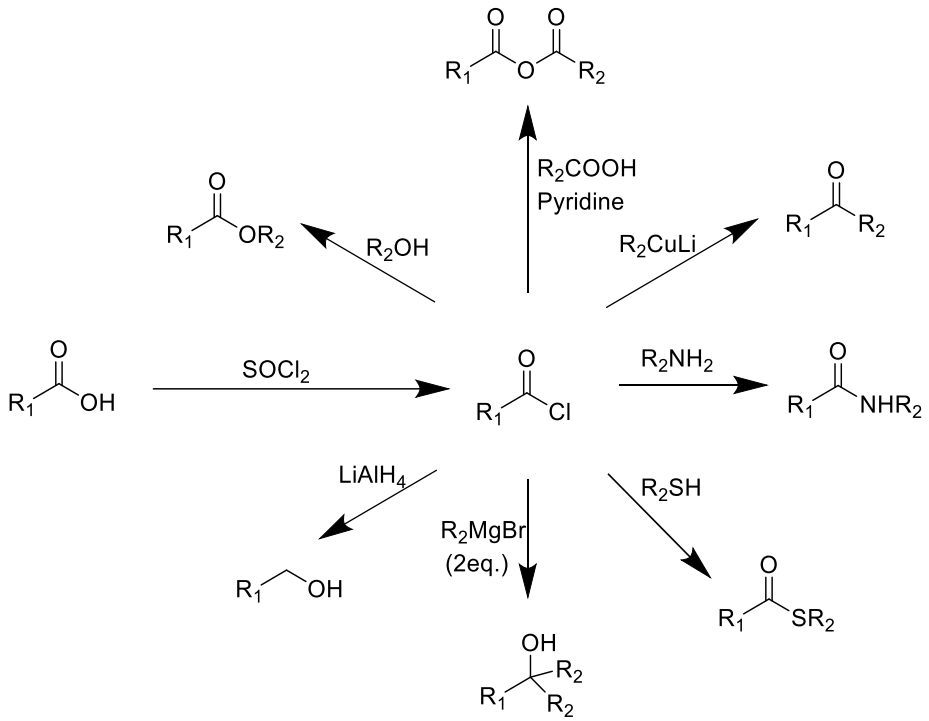


Figure 3.16: Range of compounds producible from an acyl chloride.

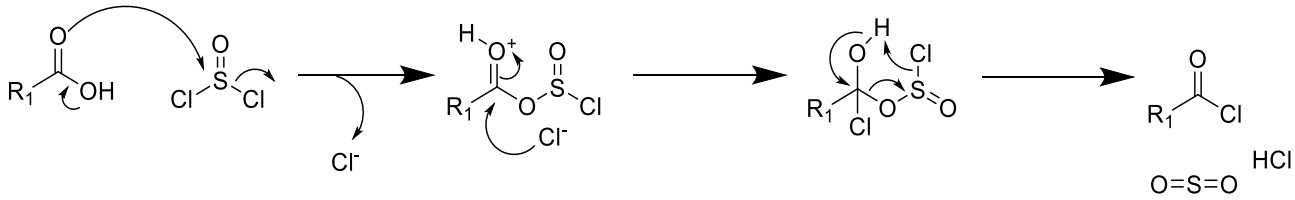


Figure 3.17: Generation of acyl chloride using thionyl chloride.

Hence, producing acyl chloride strongly increases the reactivity of the carboxylic acid. Indeed, chlorine being an excellent leaving group, any nucleophile will be able to undergo nucleophilic substitution. As such, making an acyl chloride is another way to activate the carboxylic acid and generate a better leaving group. Generation of the amide from the acyl chloride, however, is facile and only requires one equivalent of the corresponding amine moiety. As such, being very reactive, the acyl chloride is generally not isolated, and the amine is simply added to the reaction mix after a few minutes to form the amide. However, this strategy comes with similar downsides to using coupling reagents: as one equivalent of thionyl chloride is required, this reaction is not atom economical and no recycling system currently exists. In addition, while thionyl chloride is less expensive than most coupling reagents, it is also less safe to handle, being both very toxic, corrosive and violently reactive with water. For these reasons, biocatalytic alternatives are investigated, leading to new industrially applicable processes which will be discussed.

3.1.5 Amide bond formation using biocatalysis

3.1.5.1 ATP-dependent amide synthetases

The vast majority of amide synthetases (AmS) are ATP dependent. Exceptions include the GSK-developed amoxicillin synthase (discussed in section 3.1.5.2) and also include CapW, an ester-amide exchange enzyme involved in the biosynthesis of A-5003083s, a capuramycin-type antibiotic[20]. CapW requires first the activation of the carboxylic acid precursor by methylation, performed by CapS, a SAM-dependent methyltransferase. Then CapW forms the corresponding amide, as shown in Figure 3.18.

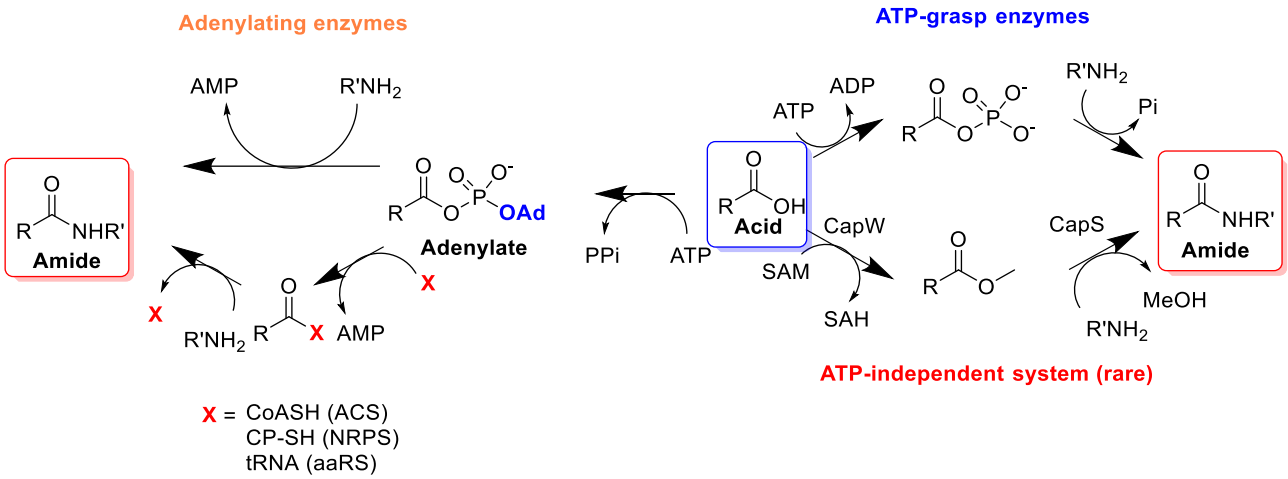


Figure 3.18: Schemes of the amide coupling reactions catalysed by adenylating enzymes, ATP-grasp enzymes and an ATP-independent system. Multiple types of AmS mechanisms exist: the adenylating enzymes, the ATP-grasp enzymes and the ATP-independent systems

The ATP-dependent enzymes can be further split into two sub-families: the adenylating enzymes and the ATP-grasp enzymes. In the ATP-grasp enzymes, an acylphosphate intermediate is formed, and ADP is released[21]. On the other hand, for the adenylating enzymes, including NS, the generation of a reactive acyl-adenylate intermediate is observed and AMP is released. The mechanism of adenylating enzymes can also differ slightly as the amine can directly attack the adenylate intermediate to make the amide (ABS), or could also proceed *via* the formation of an acyl-CoA thioester (acyl-CoA synthetase (ACS2)), a carrier protein (CP) thioester (non-ribosomal peptide synthase complexes (NRPS)) or a tRNA ester (aminoacyl-tRNA synthetase (aaRS) as shown on Figure 3.18[21]. In addition, adenylating enzymes also display structural diversity, and can be divided into four classes:

- Class I includes NRPSs, ABSs, acyl- or aryl-CoA synthetases and oxidoreductases and have a large N-terminal domain with a smaller C-terminal moiety connected via a flexible loop. The active site of these enzymes is located at the interface between these two domains[22].

An example of a Class I AmS is carbapenam synthetase, an ABS that catalyses the formation of the β-lactam ring of carbapenam-3-carboxylic acid, in the biosynthesis of clavulanic acid[23]. Its structure is shown in Figure 3.19.

- Class II includes aaRS enzymes and includes two domains: an anticodon-binding domain, for the binding of the tRNA and a central catalytic domain[22]. However, some class II enzymes have an additional amino acid-specific proofreading domain that is important for enzyme fidelity[22]. The structure of Asp-tRNA-bound yeast aspartyl-tRNA synthetase is shown in Figure 3.19.
- Class III enzymes comprise NRPS-independent siderophores and contain three domains, an N-terminal domain, a central catalytic domain as well as a C-terminal domain[24]. Examples of Class III adenylating enzymes include AcsD from *Pectobacterium chrysanthemi,* whose structure is shown in Figure 3.19[25].
- The fourth class of adenylating enzymes was a recently discovered family of enzymes: the pimeloyl-CoA synthetase (PCAS)[26]. They are composed of a small N-terminal domain and a larger C-terminal domain composed of a central antiparallel seven-stranded β-sheets surrounded by two α-helices on each sides[26]. *B. subtilis* BioW structure was determined by this group in 2017 and is shown in Figure 3.19[26].

While adenylating enzymes and ATP grasp enzymes are very important for the biosynthesis of proteins (aaRS) and natural products (other AmS), their use in biocatalysis has also been explored. While some examples will be briefly presented, a more comprehensive review, presenting further examples of AmS biocatalytic applications was published by the Campopiano group recently[21]. Examples of adenylating enzymes used in biocatalysis include McbA from *Marinactinospora thermotolerans.* This ABS enzyme was found to have a broad aryl acid specificity as it accepted a wide range of benzoic, naphthoic, indoles, quinolones and benzofurans carboxylic acids[27]. In addition, TamA from *Pseudoalteromonas tunicata*, an NRPS from the tambjamine YP1 biosynthetic pathway, was found in this group to be a broad substrate scope. This enzyme contains both an adenylation domain and an ACP domain, and it was found that the adenylation domain alone can be used as a versatile catalyst to generate a wide range of fatty *N*-acyl amides as well as a class of *N*-acyl histidine amide natural products from *Legionella pneumophilia*[28]. In addition, ATP-grasp enzymes have also been explored for uses in biocatalysis. For example, it was found that BacD, an L-aminoacid ligase found in the bacilysin, a dipeptide antibiotic synthetic pathway could display a much broader substrate scope with one single mutation[29].

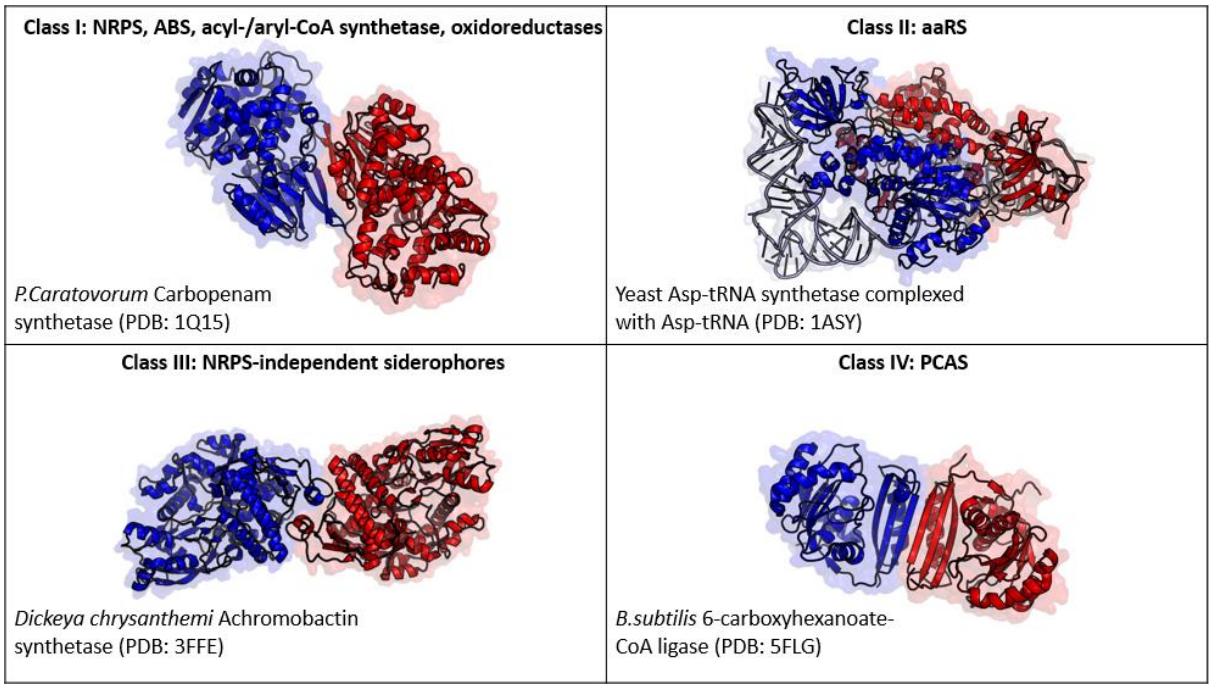


Figure 3.19: Structures of one example of each class of adenylating enzymes. Class I example is P. caratovorum carbapenem synthetase (PDB 1Q15[23]). Class II example is yeast Asp-tRNA synthetase in complex with Asp tRNA (shown in light blue and light red) (PDB 1ASY[24]). Class III example is Dickeya chrysanthemi achromobactin synthetase (PDB 3FFE[25]). Class IV example is B. subtilis pimeloyl-CoA ligase (PDB 5FLG[26])

3.1.5.2 The biocatalytic approach in action: examples of industrial uses

As discussed earlier, about 25% of small molecule drugs contain an amide function. As such, the biocatalytic route has been widely explored for the industrial-scale synthesis of commercial drugs. For example, GlaxoSmithKline (GSK) produces the β-lactam 3rd generation broad-spectrum antibiotic amoxicillin using an engineered penicillin G acylase from *E. coli*[30]. This drug is one of GSK's blockbusters and is produced at a scale of 1.1 tonnes per year[31], generating $4.9 billion revenue every year, with a CAGR of 2% until 2026[32]. The chemical and biocatalytic routes are shown in Figure 3.20. Some parameters are compared in Table 3.1.

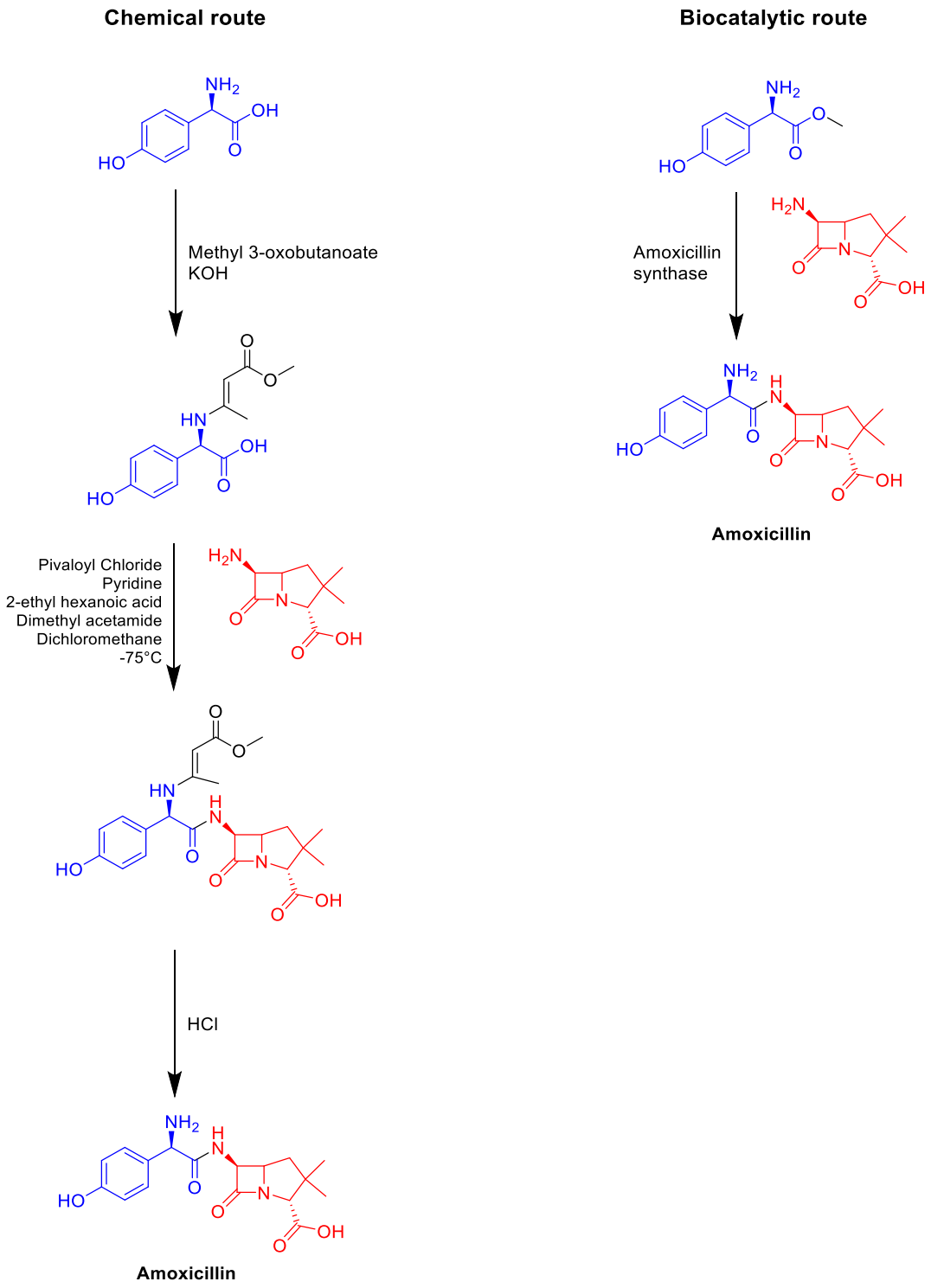


Figure 3.20: Chemical and biocatalytic syntheses of amoxicillin. The chemical route requires three steps, the use of a protecting group, extremely low temperature and the use of acid, bases and organic solvents. In contrast, the biocatalytic route only necessitates one step, mild temperature and is carried in water.

Parameter	Chemical synthesis	Biocatalytic synthesis
Number of steps	3	1
Protecting group required	Yes	No
Acids/Bases used	KOH, Pyridine, 2-ethylhexanoic acid, HCl, all in reaction	HCl and KOH used to maintain pH
Solvents used	KOH, Dimethyl acetamide, dichloromethane, HCl	Water
Temperature range	-75°C to RT	37°C
Carbon dioxide emissions	X	X – 36000 tonnes
Organic waste	X	X – 3000 tonnes

Table 3.1: Comparison of various parameters for the chemical and biocatalytic industrial syntheses of amoxicillin. X denotes an unknown number.

In addition, GSK has also been granted a patent in key territories for the lipase-catalysed synthesis of (4-isopropylpiperazin-1yl)(oxazol-5-yl)methanone, a potential phosphatidylinositol-3-OH kinase (PI3Kδ) inhibitor, a kinase involved in various immunodeficiency syndromes[33]. The chemical and biocatalytic routes to generate this compound are shown in Figure 3.21.

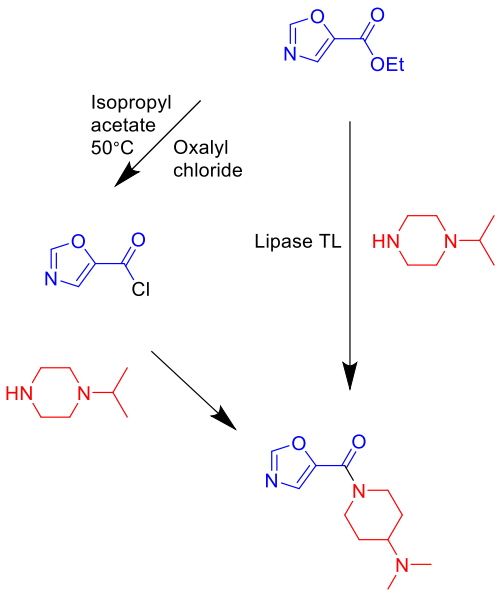


Figure 3.21: Chemical and biocatalytic syntheses of (4-isopropylpiperazin-1yl)(oxazol-5-yl)methanone. The chemical route requires the synthesis of an acyl chloride, unlike the biocatalytic route which only necissitates one step

Hence, the first example shows an example where an enzyme replaced a coupling reagent while the second one shows a case where a biocatalyst-driven strategy substituted a synthetic route involving an acyl chloride. While the economic and environmental benefits of biocatalysis were

already discussed, this shows that enzymes can also provide an efficient alternative for the industrial production of amides.

In addition, as developed earlier, most amide-producing enzymes are ATP-dependent. These, however, have a major downside for their industrial use as they require one equivalent of ATP. As such, similarly to chemical approaches, these enzymes would require a stoichiometric amount of ATP, which would, just like a coupling reagent, be used to activate the carboxylic acid. However, this consideration is alleviated by the fact that, unlike for coupling reagents, ATP recycling systems have been developed. An example of such recycling systems is the use of polyphosphate, phosphotransferase (PPT) and adenylate kinase (AdK)[34]. Upon consumption of the ATP by the AmS, generating adenosine monophosphate (AMP) or adenosine diphosphate (ADP), AdK will be able to transfer phosphate groups from one or two molecules of ADP[35]. The phosphate donor, transformed into AMP, will be regenerated into ADP by PPT, which will use cheap polyphosphates such as polyphosphoric acid as a phosphate source[34]. Examples of the successful use of this recycling system include the regeneration of ATP in the carboxylic acid reductases-catalysed transformation of acids into ketones[36] by the group of Nicholas Turner or the enzymatic generation of glutathione by GSH synthase[37]. The scheme of this ATP recycling system is shown in Figure 3.22.

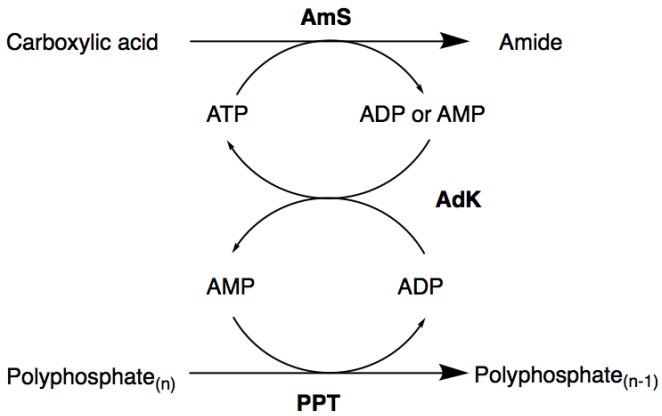


Figure 3.22: Phosphotransferase/Adenylate kinase recycling system scheme: The ADP or AMP generated from the transformation of the carboxylic acid to the amide is recycled by the PPT/AdK system, using a polyphosphate as the phosphate donor.

3.1.6 Making nitriles from carboxylic acids.

3.1.6.1 The direct route

In 1872, Edmund Letts published a method for the synthesis of aromatic nitriles directly from the corresponding carboxylic acids using an excess of potassium thiocyanate and heat with yields of up to 50%[38]. Kekule has proposed a mechanism for this transformation: the potassium thiocyanate will first

deprotonate the acid. Then, the benzoic acid derivative will undergo thermal decomposition while heated and release carbon dioxide and a benzene anion[39]. The benzene anion will then undergo nucleophilic substitution on potassium thiocyanate, forming benzonitrile and potassium hydrosulfide[38]. The reaction scheme and mechanism are shown in Figure 3.23

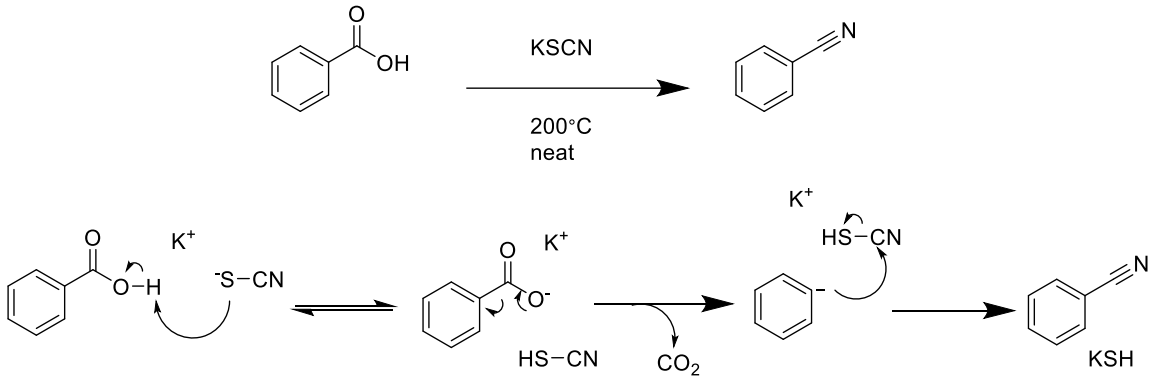


Figure 3.23: The Letts reaction scheme and proposed mechanism: potassium thiocyanate deprotonates benzoic acid, which will release carbon dioxide through decomposition, generating a benzene anion. The anion will undergo nucleophilic substitution, generating benzonitrile and potassium hydrosulfate.

This reaction, however, is not very efficient with only a 50% yield. In addition, as the protocol involves hard heating of benzoic acid and potassium thiocyanate (both solid at room temperature), above their melting points in neat conditions. Upon melting, the reagents will form a liquid bilayer for the reaction to occur. In addition, this reaction leads to the evolution of potassium hydrosulfide, a toxic gas. This reaction was trialled in the laboratory. However, instead of heating the solids, the reaction was refluxed at 200°C in anhydrous toluene. However, no benzonitrile was formed.

Other direct transformations from carboxylic acid to nitrile were found to be difficult and require high temperatures (ammonia and silica at 500°C[40], ammonium hydroxide at 600°C[41] or ammonia in continuous flow at 360°C[42]), although microwave irradiation could provide a promising alternative as Juncai and colleagues were able to convert dodecanoic acid into the corresponding nitrile with a 79% yield through microwave irradiation, using aluminium oxide, aminosulfonic acid and urea[43].

3.1.6.2 The amide route

A more convenient route to generate a nitrile from a carboxylic acid is to first generate the amide and then dehydrate it. The amide can be generated as described in Section 3.1.5. Various conditions exist for the amide dehydration, amongst which the use of thionyl chloride[44], palladium chloride[44], phosphorus pentoxide[44] or conditions close to those of the Appel reaction (oxalyl chloride, triphenylphosphine oxide and triethylamine)[45]. The proposed mechanism of this dehydration reaction

for thionyl chloride is shown in Figure 3.24. It involves the attack of the amide on thionyl chloride. After rearrangement, the amide is deprotonated, and the leaving group is removed.

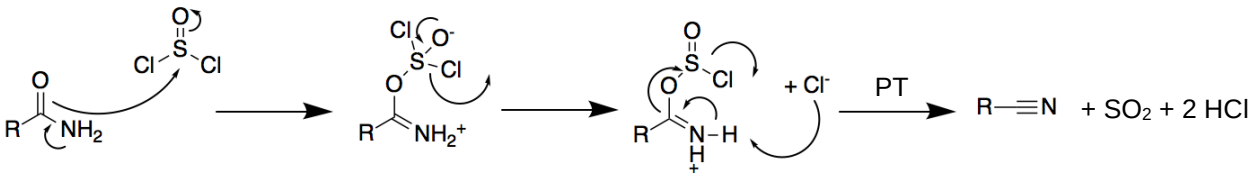


Figure 3.24: Mechanism of nitrile synthesis from a primary amide using thionyl chloride: the amide attacks thionyl chloride, inducing a rearrangement which leads to the liberation of a chloride anion. This anion will deprotonate the amide and induce the removal of the leaving group to afford the nitrile.

Interestingly, except for the NS, no enzyme was shown to catalyse amide dehydration, emphasizing the unusual nature of NS enzymes which will be presented in this project.

3.1.7 Aims

The aims of this project are:

- To express, purify and characterise NS enzymes
- To design an assay to monitor efficiently and precisely the NS reaction
- To optimise the NS reaction
- To carry out a structural study of the NSs and, more specifically, to obtain a ligand-bound structure of one of the NS enzymes
- To use this ligand-bound structure to carry out rational engineering of the enzyme in order to generate a novel amide synthetase (BsQueC eAmS K163A, R204A).
- To characterise the kinetics of the BsQueC eAmS K163A, R204A and use this to prepare ADG from CDG on a synthetic scale.

3.2 Generation of the biocatalysts

To commence this study, two recombinant NSs were prepared: *B. subtilis* QueC (BsQueC; Uniprot: O31675) and *S. rimosus* ToyM (SrToyM; Uniprot: B6CWJ6). Both enzymes were cloned in pET28a, a plasmid previously introduced in Chapter 3, so the recombinant form of the enzymes will contain an N-terminal hexahistidine tag. Both enzymes were expressed in *E. coli* and purified by IMAC and SEC. Plasmid maps are available in the Appendix.

3.2.1 Generation of BsQueC

The BsQueC (Mwt: 26535 Da) cloning and expression optimisation were carried out by Mari Nelson, a previous student in the Campopiano group. The expression tests SDS-PAGE gel is shown in Figure 3.25.

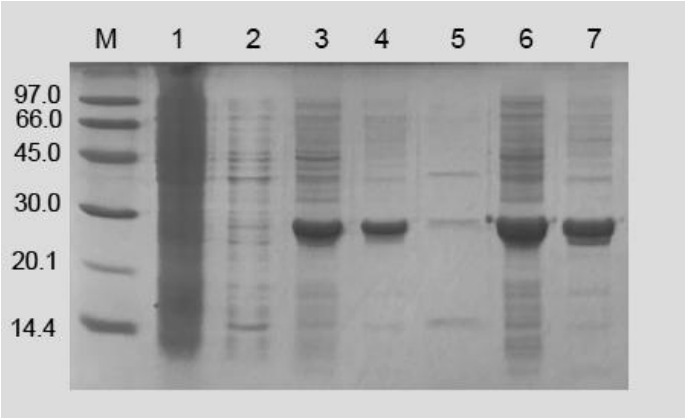


Figure 3.25: 12% SDS-PAGE gel for BsQueC expression tests. M: low molecular weight ladder; 1: Pre-IPTG; 2: 0.1 mM IPTG, 30°C, 3h; 3: 0.1 mM IPTG, 30°C, 5h; 4: 0.1 mM IPTG, 18°C, 18h; 5: 0.5 mM IPTG, 30°C, 3h; 6: 0.5 mM IPTG, 30°C, 5h; 7: 0.5 mM IPTG, 18°C, 18h. 0.1 mM $ZnSO_4$ was also added in all tested conditions. Only soluble fractions are shown. A large band at 26 kDa, thought to correspond to BsQueC was observed in conditions 3, 4, 6 and 7. Conditions 6 were concluded to be the best. Figure generated by Mari Nelson.

Hence, the optimised expression protocol involves expressing the *B. subtilis QueC* gene in *E. coli* BL21 (DE3) strain for 5 hours at 30°C after induction with 0.5 mM IPTG and in the presence of 0.1 mM $ZnSO_4$, as a zinc source. The enzyme was then purified by nickel affinity chromatography using a 40-500 mM imidazole gradient. The enzyme eluted at 220 mM imidazole and SDS-PAGE analysis showed traces of a few higher molecular weight proteins. The protein was then purified by SEC using a Superdex S200 HR column. The protein eluted at 66 mL, suggesting, using the column calibration curve shown in the Appendix, that BsQueC eluted as a hexamer (Determined Mwt of the complex from S200 calibration curve: 164 kDa; Mwt of BsQueC hexamer: 159 kDa). This oligomeric state was expected as per the literature[15]. The SEC chromatogram is shown in Figure 3.26.

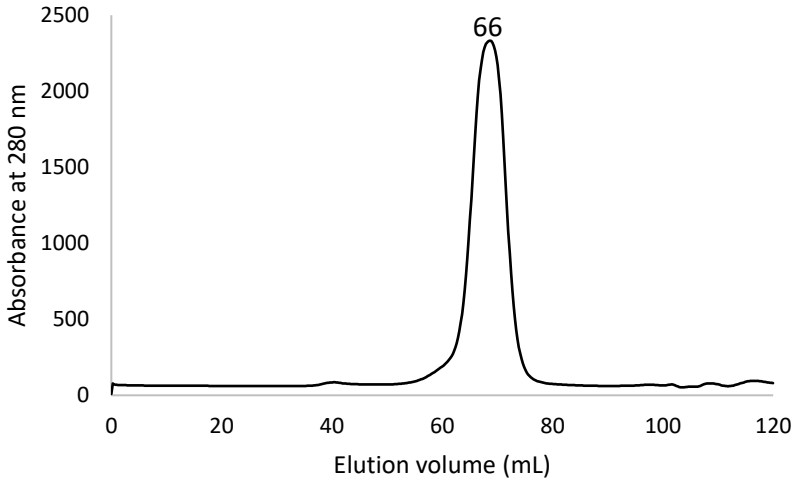


Figure 3.26: Superdex S200 size-exclusion chromatogram of BsQueC. A single peak corresponding to BsQueC is observed at an elution volume of 66 mL. From the column calibration curve, this corresponds to a Mwt of 158 kDa, suggesting that BsQueC eluted as an hexamer (Mwt 159kDa)

The BsQueC SDS-PAGE analysis of a typical purification is shown in Figure 3.27. The purified enzyme was also characterised by ESI-MS, following SEC. The ESI-MS spectrum is shown in Figure 3.28.

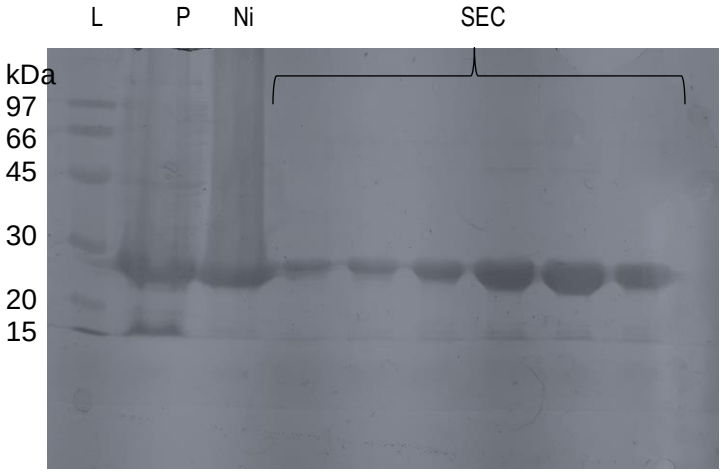


Figure 3.27: 12 % SDS-PAGE gel of a hexahistidine tagged BsQueC purification steps.
L: Low molecular weight ladder, P: Cell pellets, Ni: Sample after immobilised nickel affinity chromatography. SEC: Samples after size-exclusion chromatography

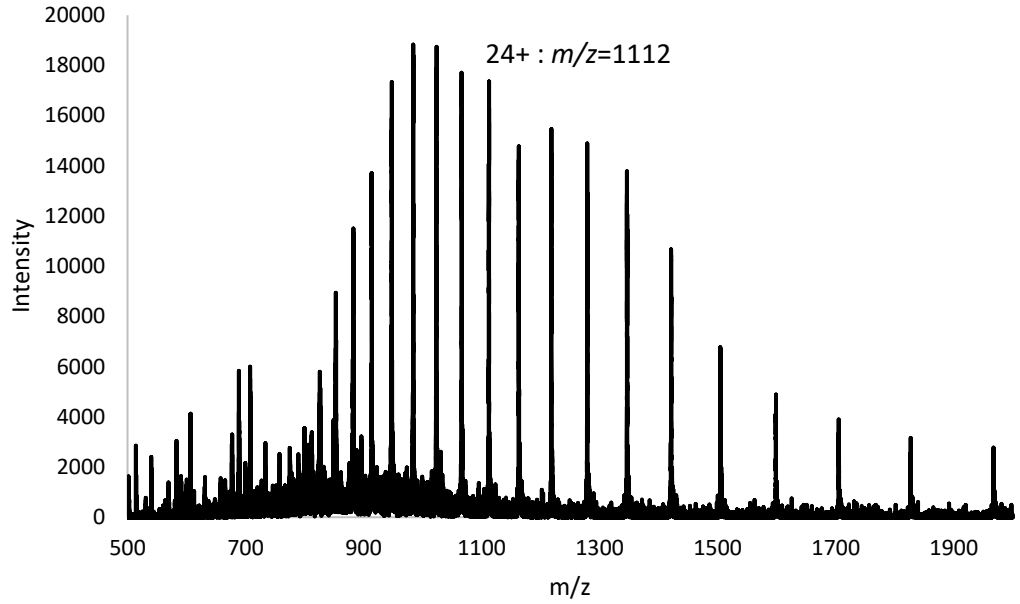


Figure 3.28: BsQueC ESI-MS spectrum. The BsQueC experimental mass was found to be 26535.89 ± 0.73 Da and in agreement with its theoretical mass (ExPasy): 26535.31 Da.

Hence, as shown in Figure 3.3, while immobilized nickel affinity chromatography recovered the protein with an estimated purity of 80%, the purity after SEC was increased to >95%. The experimentally-determined mass was 26535.89 ± 0.73 Da, in good agreement with the predicted mass (26535.31 Da). Yields of 20-25 mg of pure protein/L of culture (LB) were routinely obtained.

3.2.2 Generation of *SrToyM*

The *S. rimosus toyM* codon-optimised gene was bought from GenScript and shipped in a pUC57 plasmid. The gene was then cloned into a pET28a plasmid, so the recombinant protein will feature a N-terminal hexahistidine tag, and expression optimisation was carried out, as shown in Figure 3.29.

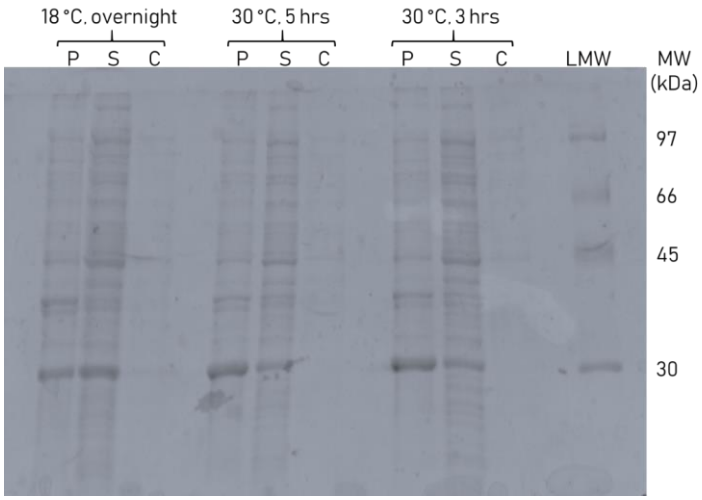


Figure 3.29: Expression tests for SrToyM. In all the experiments, the cells were induced with 0.1 M IPTG and 0.1 M $ZnSO_4$. P: insoluble fraction, S: soluble fraction, C: Control (uninduced). Overnight expression at 18°C was found to be the best conditions.

As a band of good intensity was observed near the predicted 30 kDa region on the gel shown on Figure 3.5, it was concluded that best expression was obtained after expressing the *toyM* gene in *E. coli* BL21 (DE3) strain for 18 hours at 18°C after induction with 0.1 mM IPTG and in the presence of 0.1 mM $ZnSO_4$, as a zinc source. SrToyM large scale expression (5 L) was carried out. As for BsQueC, SrToyM was purified by nickel affinity chromatography using a 40-500 mM imidazole gradient. The enzyme eluted at 300 mM imidazole and SDS-PAGE analysis showed traces of a few higher molecular weight proteins. The protein was then purified by SEC using a Superdex S200 column. The protein eluted at 61 mL, suggesting that SrToyM also eluted as an octamer (Determined Mwt of the complex from S200 calibration curve: 224 kDa; Mwt of SrToyM octamer: 220 kDa). The SEC chromatogram is shown in Figure 3.30, along with an SDS-PAGE gel of a routine SrToyM purification.

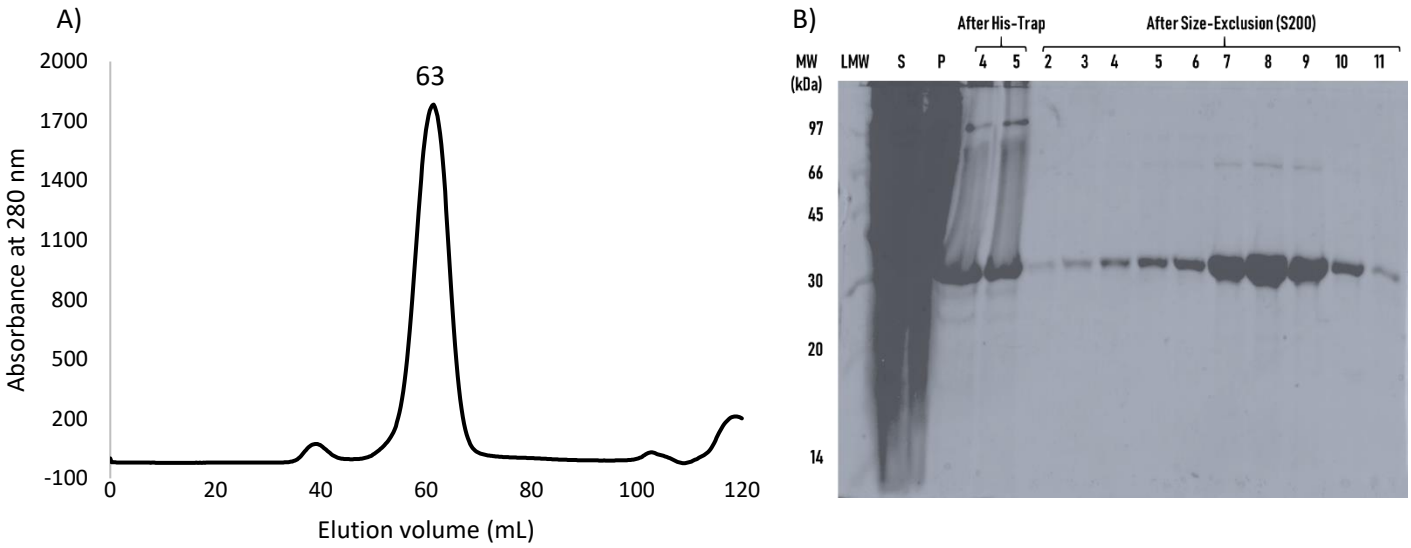


Figure 3.30: A): Superdex S200 size-exclusion chromatogram of SrToyM A single peak corresponding to the enzyme is observed at an elution volume of 61 mL.
B): 12% SDS-PAGE gel of a hexahistidine tagged SrToyM purification steps. S: Cell-free lysate; P: Cell pellets (insoluble).

As shown on the SDS-PAGE analysis, enzyme purity superior to 95% was achieved after the SEC. Yields of 10-15 mg of pure protein/L of culture (LB) were routinely obtained. As for the other proteins characterised, SrToyM was also analysed by ESI-MS. The ESI-MS spectrum is shown in Figure 3.31.

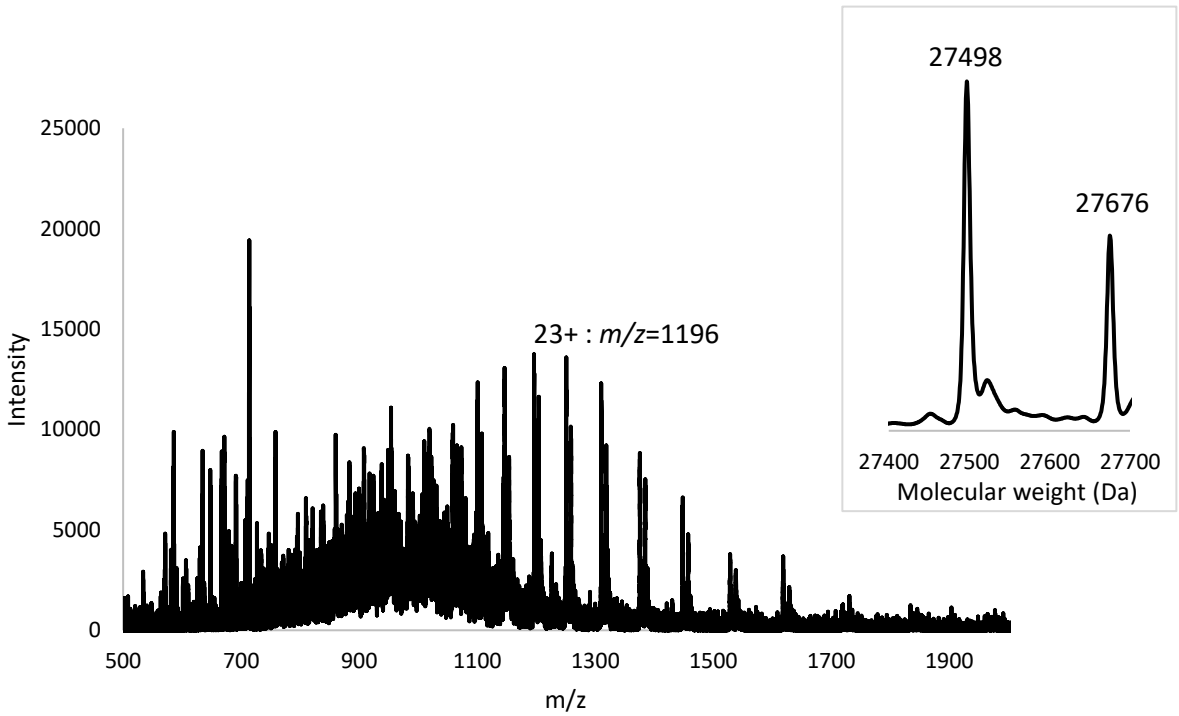


Figure 3.31: SrToyM ESI-MS spectrum. The SrToyM theoretical mass (ExPasy) is 27497.71 Da. SrToyM actual mass was 27498.34 ± 1.12 Da. The two sets of peaks correspond to apo-ToyM and to glycophosphorylated SrToyM and have a mass difference of 178 Da.

The experimentally-determined mass was 27498.34 ± 1.12 Da, in good agreement with the predicted mass (27497.11 Da). A second peak was observed at 27676 Da. This peak corresponds to the glycophosphorylated form, typically observed with recombinant proteins expressed in *E. coli.*

3.3 Generation of the substrates and products

3.3.1 Syntheses of CDG and $PreQ_0$

CDG and $PreQ_0$ were synthetised from commercially-available products following a route adapted from Nelp *et al.* work[13] shown in Figure 3.32.

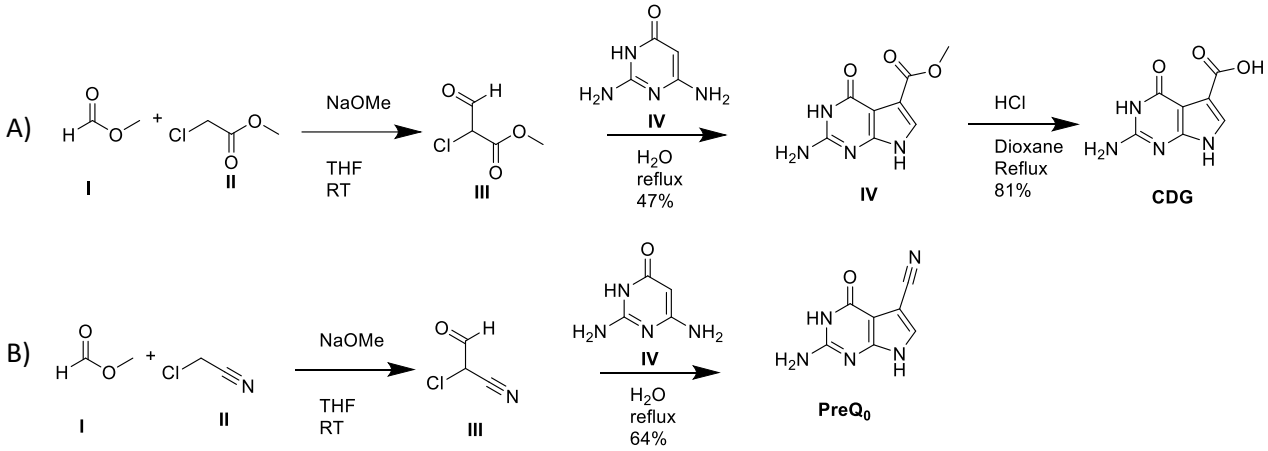


Figure 3.32: A) CDG and B) $PreQ_0$ synthetic schemes.

Hence, the first step of these syntheses is the Claisen condensation of methyl formate (**I**) with chloroacetonitrile or methyl chloroacetate (**II**) to form a chloroaldehyde (**III**) which will be able to undergo a Knorr reaction with 2,4-diamino-6-pyrimidinone (**IV**) to generate the pyrrole moiety of the deazaguanine ring (**V**). The Knorr reaction is commonly used to generate substituted pyrroles. It starts with the nucleophilic attack of one of the pyrimidinone free amines on the carbonyl. After the elimination of water, a double bond will be formed. Through rearrangement, the π-electrons of one of the pyrimidinone moiety will undergo nucleophilic substitution, eliminating Cl^- and closing the pyrrole ring, forming the deazaguanine core. The proposed mechanism of this reaction is shown in Figure 3.33.

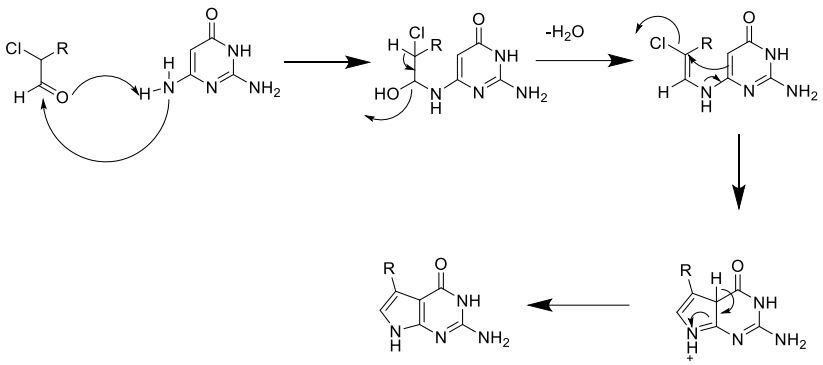
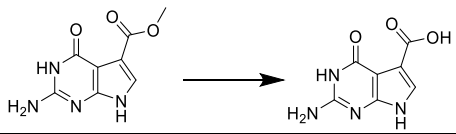


Figure 3.33: Knorr pyrrole synthesis mechanism.

After carrying out these two steps, that can be done in one pot, without isolating the chloroaldehyde, both $PreQ_0$ and CDG-methyl ester was isolated with yields in the region of 70-75% and purities of >95% (by HPLC and NMR). CDG was then generated through the hydrolysis of CDG methyl ester. Some optimisation, in both basic, acidic and biocatalytic conditions was carried as shown in Table 3.2.



	Reagent	Solvent	Temperature (°C)	Conversion after 5 hours (%)	Yield (%)	Remarks
Basic conditions	NaOH (3 M)	H_2O	130	100	Not isolated	Mixture of by-products
	NaOH (3 M)	H_2O/DMSO (1:1)	RT	0	Not isolated	
	NaOH (1 M)	H_2O	130	57	Not isolated	Mixture of by-products
Biocatalysis	BioH (0.12 mM)	GFC buffer	37	0	Not isolated	
Acidic conditions	HCl (1 M)	H_2O	130	13	Not isolated	
	HCl (6 M)	H_2O/DMSO (1:1)	RT	0	Not isolated	
	HCl (6 M)	H_2O/THF (1:1)	130	100	264	Formation of 4-chlorobutanol
	HCl (6 M)	H_2O/Dioxane (1:1)	130	100	76	Purity: 95+%

Table 3.2: Optimisation of the CDG-methyl ester hydrolysis reactions. Best conditions are acidic hydrolysis with HCl (6 M) in H_2O/dioxane (1:1) at 130°C.

It was found that basic hydrolysis was not a suitable method to generate CDG as the methyl ester was fully converted into a mixture of by-products upon carrying the reaction with concentrated sodium hydroxide and reflux conditions. Upon lowering the temperature to RT, in this case, no

reaction was observed. It was also observed that lowering the concentration of sodium hydroxide did not provide any improvement as, in this case, conversion of the starting material was incomplete, and the desired product was not observed. Acidic hydrolysis was found to be much more efficient: full conversion of the methyl ester to the acid was observed when using 6M HCl in reflux conditions. While this was first tested in water and THF, it was observed that THF was undergoing a ring-opening reaction with HCl to form 4-chlorobutanol. This compound was very hard to remove from the mix, but this problem was fixed by using 1,4-dioxane instead, which does not undergo any ring-opening reaction. Hence, CDG was isolated with a 76% yield and 95+% purity. Moreover, *E. coli* BioH (EcBioH), being a methyl esterase (see Chapter 3), was also tested but did not exhibit any activity for the substrate. As robust conditions were found for this reaction, the biocatalytic route was not further explored.

3.3.2 Synthesis of ADG

Surprisingly, the corresponding primary amide was much more challenging to obtain. Through searching on the Reaxys database, four synthetic routes were available. Two of which were described in the Nelp *et al.* paper[13], and involve the use of enzymes (ToyJ nitrile hydratase from $PreQ_0$ and SrToyM from CDG, although ADG was not isolated in the SrToyM route as it was readily converted into $PreQ_0$). This synthetic route is shown in Figure 3.8. The two other routes used either $PreQ_0$[46] or CDG-methyl ester[47] and were tried as shown in Table 3.3.

Reference	Starting material	Conditions	Conversion after 24 h (%)	Remarks
[2]	$PreQ_0$	Sodium percarbonate (1.5 eq), potassium hydroxide (0.5 M), RT	3	
[3]	CDG-methyl ester	Ammonia (excess), methanol (solvent), 120°C	24	CDG observed

Table 3.3: Trial of literature conditions to generate ADG.

As no satisfactory results were obtained through reproducing literature conditions, four different approaches were explored for generating ADG: peptide coupling and acyl chloride formation on CDG, hydrolysis of $PreQ_0$ nitrile and the direct conversion of CDG-methyl ester into ADG. Findings are shown in Table 3.4.

Approach	Reagent	Temp (°C)	Conversion (%) (time in brackets)	Yield (%)	Remarks
Coupling reagent **SM: CDG**	HATU (1.5 eq) + ammonia (2.5 eq)	RT	0 (24 hours)	Not isolated	
	HATU (5 eq) + ammonia (2.5 eq)	130	0 (24 hours)	Not isolated	
	DCC (1.5 eq) + ammonia (2.5 eq)	RT	0 (24 hours)	Not isolated	
	DCC (5 eq) + ammonia (2.5 eq)	130	0 (24 hours)	Not isolated	
	TOTT (1.5 eq) + ammonia (2.5 eq)	RT	0 (24 hours)	Not isolated	
	TOTT (5 eq) + ammonia (2.5 eq)	130	0 (24 hours)	Not isolated	
Formation of the acyl chloride **SM: CDG**	1) Thionyl chloride (1.5 eq) 2) ammonia (excess)[1]	RT	0 (48 hours)	Not isolated	
	1) Thionyl chloride (excess) 2) ammonia (excess)[2]	RT	0 (48 hours)	Not isolated	
	1) Oxalyl chloride (5 eq) + DMF (cat. amount) 2) ammonia (excess)[1]	RT	0 (48 hours)	Not isolated	
	1) Thionyl chloride (1.5 eq) 2) ammonia (excess)[1]	130	0 (24 hours)	Not isolated	
Hydrolysis of $PreQ_0$ nitrile **SM: $PreQ_0$**	Sodium percarbonate (1.5 eq), potassium hydroxide (1 M)	RT	8 (24 hours)	Not isolated	
	Sodium percarbonate (1.5 eq), potassium hydroxide (1 M)	130	72 (24 hours)	41	Purity (NMR) = 58%, traces of unreacted starting material and CDG
Direct conversion CDG-Methyl ester to ADG **SM: CDG-Me ester**	Magnesium nitride (5 eq) in a sealed tube	80	20 (24 hours)	7	Purity (NMR) = 27%. Mix of CDG Me-ester, CDG and $PreQ_0$
	Magnesium nitride (15 eq) in a sealed tube	80	22 (24 hours)	Not isolated	

Table 3.4: Conditions screen for the generation of ADG. Best conditions are the hydrolysis of $PreQ_0$ nitrile using sodium percarbonate (1.5 eq), potassium hydroxide (1 M) at 130°C.
SM: starting material; DCC: dicyclohexyl carbodiimide; TOTT: 2-(1-Oxy-pyridin-2-yl)-1,1,3,3-tetramethylisothiouronium tetrafluoroborate.
[1]: 10 mL of ammonia (13.4 M) solution were added after 5 hours of reaction with thionyl/oxalyl chloride.
[2]: 5 mL thionyl chloride was added (neat). After 5 hours of reaction, 15 mL ammonia (13.4 M) solution were added.

The best found conditions still resulted in low yields. As ADG (as well as CDG and $PreQ_0$) is insoluble in organic solvents and only soluble in DMSO and highly concentrated strong bases (500 mM KOH), no conditions for the separation of ADG, $PreQ_0$ and CDG by flash column chromatography were found,

hence the low achieved purity. While this could still be used to determine the ADG HPLC retention time for assay development, it was also concluded that biocatalysis should be a better route for the high yield and purity generation of ADG.

3.4 Towards a nitrile synthetase assay

3.4.1 Development of an HPLC assay

The first assay developed for this study was an HPLC assay, separating all the reagents and products (CDG, ADG, $PreQ_0$, MgATP, and MgAMP). As such, all the reagents and products were tested through running each compound on a 0-95% acetonitrile gradient at 254 nm on a Phenomenex Luna 5u C18 reverse phase HPLC column to assess at which acetonitrile concentration each compound elutes. It was found that all MgATP and MgAMP eluted at 0% acetonitrile while CDG, ADG and $PreQ_0$ eluted at 15-20%. This separation was further optimised, through modifying the acetonitrile gradient and, in the finalised protocol, two region could be observed in the chromatogram, one containing the MgATP and MgAMP peaks (3-5 min) and one containing CDG, ADG and $PreQ_0$ peaks (18-21 min). A chromatogram featuring all the compounds observed in the assay is shown in Figure 3.34.

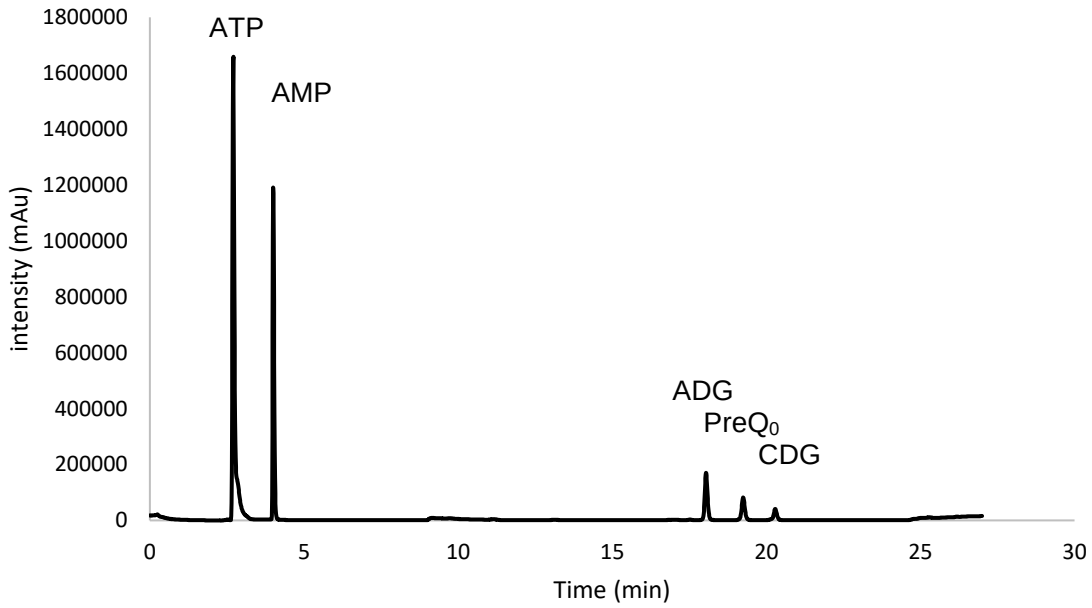


Figure 3.34: HPLC chromatogram of MgATP, MgAMP, ADG, $PreQ_0$ and CDG. Mobile phase: 0-20% acetonitrile + 0.1% TFA in water + 0.1% TFA

The region of the HPLC chromatogram featuring ADG, $PreQ_0$ and CDG is shown in Figure 3.35:

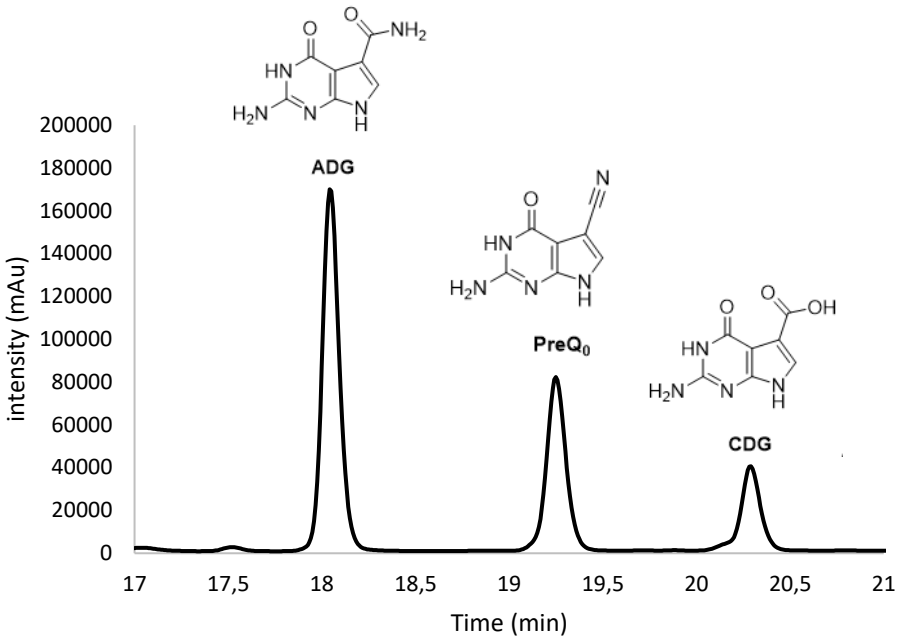


Figure 3.35: HPLC chromatogram of ADG, $PreQ_0$ and CDG. Mobile phase: 15-20% acetonitrile + 0.1% TFA in water + 0.1% TFA.

Hence, as efficient separation of the CDG starting material and the two products was achieved, calibration curves (shown in appendix) were obtained for CDG and $PreQ_0$ to determine their extinction coefficients at 254 nm. A summary of the HPLC retention times of each compound is shown in Table 3.5. This method being only 27 mins long, it represents an improvement to Nelp *et al.* HPLC protocol, which was 40 mins long[13].

Compound	Retention time (min)	Extinction coefficient (L/mol/cm) (254 nm)
MgATP	2.7	NC
MgADP	2.8	NC
MgAMP	4.0	NC
ADG	18.1	NC
$PreQ_0$	19.3	487430
CDG	20.4	207780

Table 3.5: HPLC retention times of each reagent and products of the NS reaction.

3.4.2 Can an efficient HT coupled assay be found ?

The first potential HT assay which was considered was to couple an NS with the next enzyme in the queuosine biosynthetic pathway, QueF from *B. subtilis* (BsQueF). BsQueF is an NADPH-dependent nitrile reductase which converts $PreQ_0$ nitrile into the corresponding amine[48]. As NADPH, a UV-active

compound (340 nm) is converted into $NADP^+$, which does not absorb UV, NADPH disappearance can be easily monitored at 340 nm. Two equivalents of NADPH are required for this reaction. This reduction is thought to involve the nucleophilic attack of a cysteine residue in QueF catalytic triad on $PreQ_0$ nitrile to form a covalent thioimide intermediate. Then, one equivalent of NADPH will deliver a hydride to form a thiohemiaminal intermediate which will be broken down into an imine by the hydride delivered by the second NADPH equivalent[49]. The scheme of this assay and the mechanism of BsQueF reduction are shown in Figure 3.36.

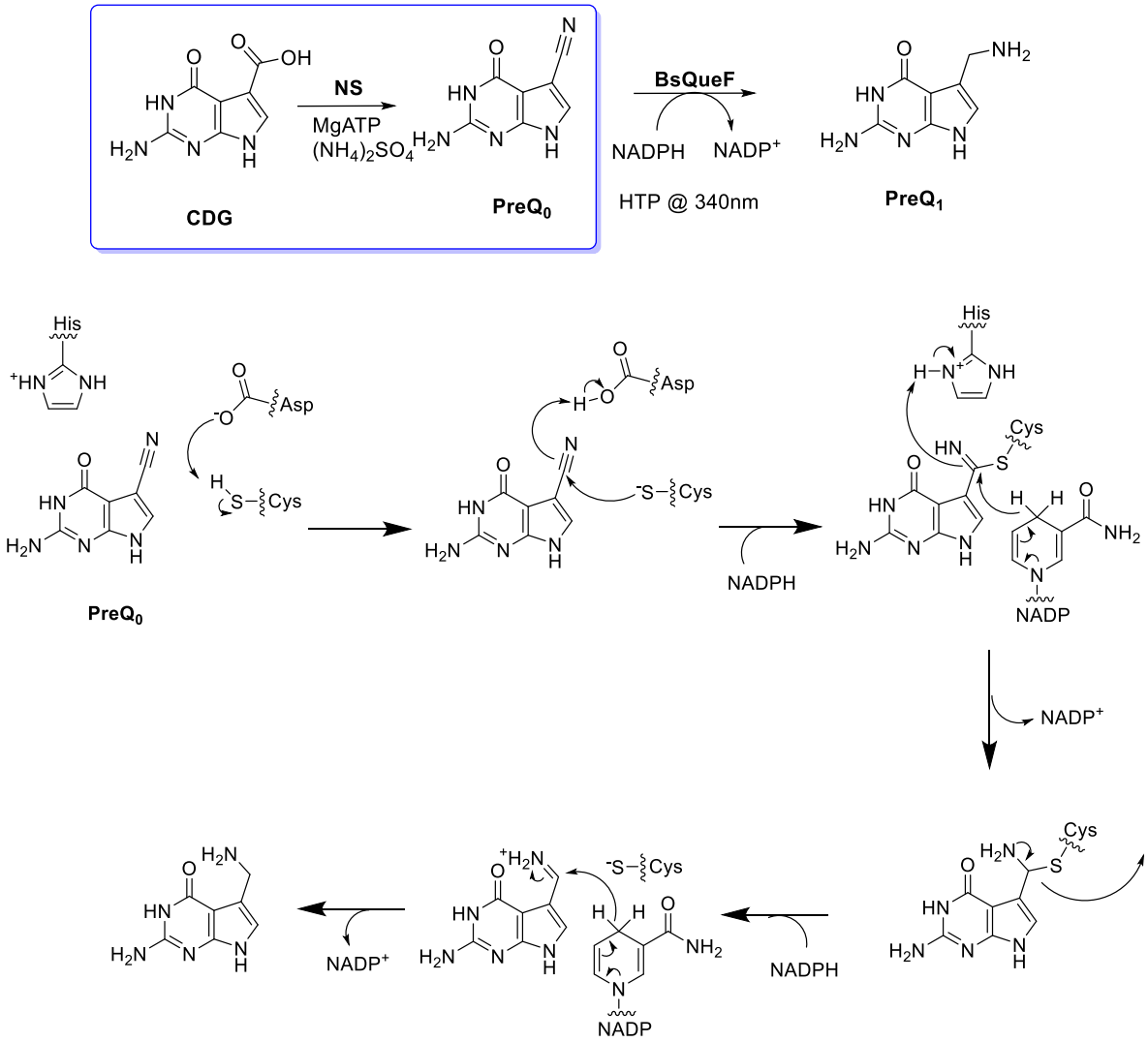


Figure 3.36: Coupled NS-QueF assay scheme and mechanism of the QueF-catalysed reduction of $PreQ_0$.

BsQueF (Uniprot: O31678) was expressed, purified (IMAC and Superdex S200 SEC) and characterised following the same procedure described for SrToyM with the exception that no zinc sulfate was added during protein expression. BsQueF SDS-PAGE gel and ESI-MS spectrum are shown

in Figure 3.37. It was found that the mass observed by ESI-MS (20440.08 ± 0.35 Da) is in agreement with the predicted mass (20440.30 Da)

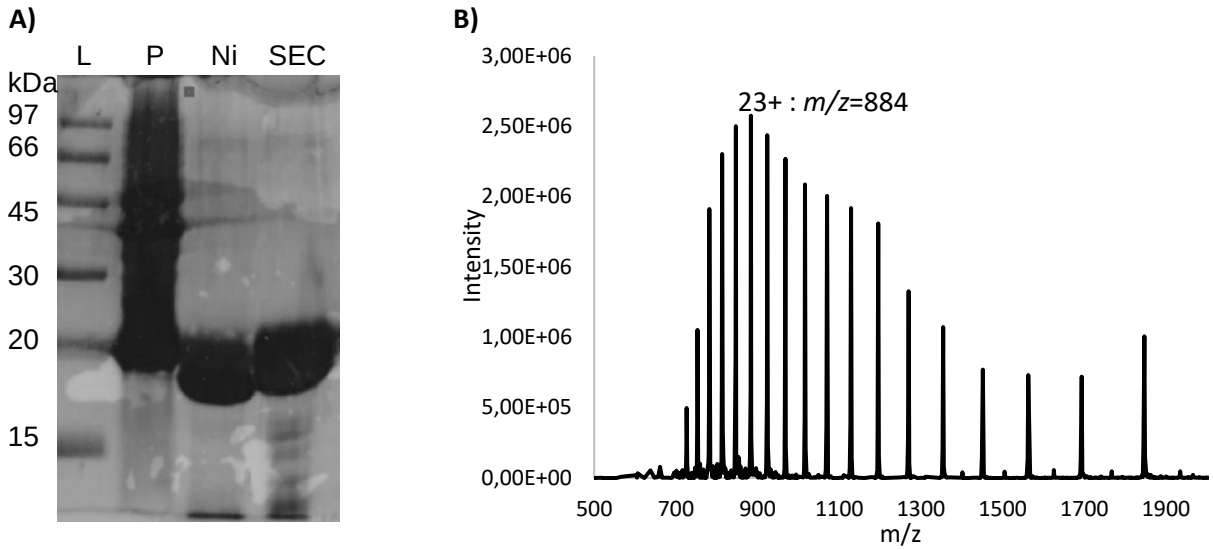


Figure 3.37: A): BsQueF S200 chromatogram; B): BsQueF ESI-MS spectrum. Theoretical mass (ExPasy): 20440.30 Da; Actual mass: 20440.08 ± 0.35 Da.

First, BsQueF was assayed with $PreQ_0$ as substrate, and two equivalents of NADPH and its kinetic parameters were determined. The Michaelis-Menten plot is shown in Figure 3.38, and the determined kinetic parameters are shown in Table 3.6.

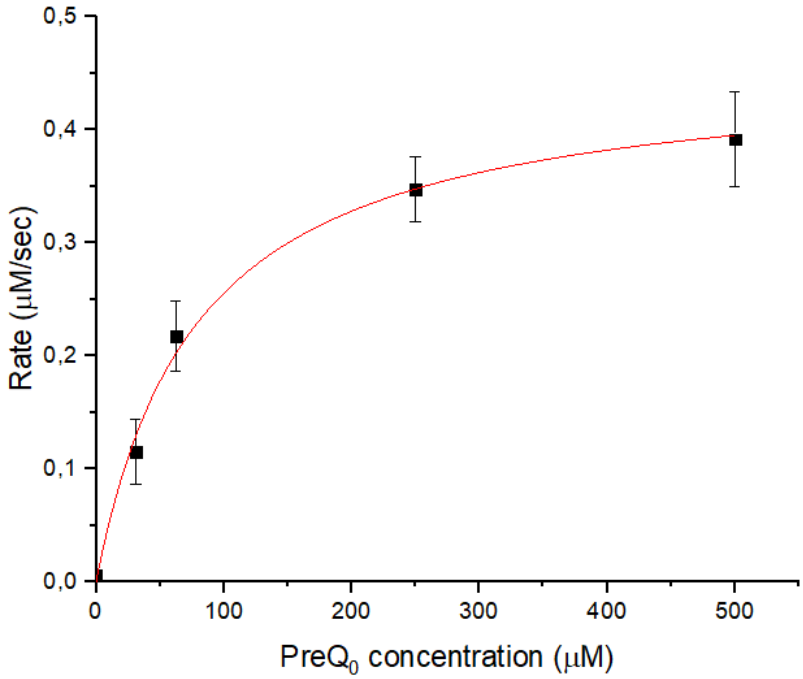


Figure 3.38: BsQueF Michaelis-Menton plot for $PreQ_0$

Michaelis constant	K_M	78.9 ± 9.1 µM
Turnover number	k_{cat}	0.0093 ± 0.0038 s^{-1}
Specificity constant	k_{cat}/K_M	1.2E-4 ± 1.4E-5 $\mu M^{-1}.s^{-1}$

Table 3.6: BsQueF kinetic parameters for $PreQ_0$ reduction.

This enzyme catalyses the reaction at a low rate as it displays a k_{cat} of 0.0093 s^{-1} or 0.558 min^{-1}. This is consistent with literature results as a k_{cat} of 0.66 min^{-1} was previously reported[50]. However, while coupling NSs with BsQueF could be a good assay to detect $PreQ_0$, ADG cannot be detected with this method. As the aim of this study is to split the CDG to ADG and the ADG to $PreQ_0$ reactions in order to engineer an AmS from an NS, a robust assay for this project would be an assay where both products could be detected. In addition, while some phosphate detection assays such as the methylthioguanosine assay[51] or the malachite green assay[52] were also investigated, these assays were not optimised either because these would not be able to differentiate the formation of ADG and $PreQ_0$ from CDG. As such, these high throughput assays are not optimal for this study. It was therefore concluded that the HPLC assay would be the best assay to quantify directly the relative and absolute amounts of CDG, ADG and $PreQ_0$. A summary of these assays is shown in Table 3.7.

	HPLC assay	NS-QueF coupled assay	Phosphate detection assays
Compound(s) detected	CDG, ADG and $PreQ_0$	$PreQ_0$	Phosphate
Throughput	Low-medium	High	High
Ability to detect both ADG and $PreQ_0$	Yes	No (ADG not detected)	Yes
Ability to distinguish ADG and $PreQ_0$	Yes	Yes	No

Table 3.7: Summary of the properties of the investigated assays.

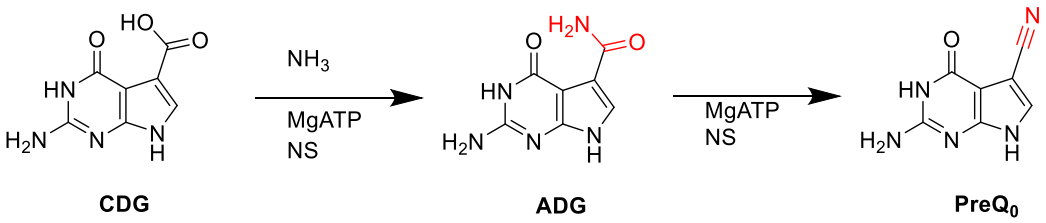
3.5 Assaying NSs

3.5.1 Optimisation of reaction conditions

Several conditions (the enzyme and CDG loadings and the number of equivalents of MgATP) were screened using BsQueC in order to identify efficient reaction conditions for the NS-catalysed biotransformation. The best conditions obtained after this screen were also used for SrToyM to allow direct comparison between the enzymes.

3.5.1.1 Enzyme loading

First, BsQueC loading was investigated for the transformation of CDG to ADG and $PreQ_0$. The conversion of CDG was assessed after 5 hours, as shown in Table 3.8.



CDG **ADG** **$PreQ_0$**

Enzyme loading (mg/mL)	CDG (%)	ADG (%)	$PreQ_0$ (%)
2	4	34	62
1	5	60	35
0.25	7	89	4
0	100	0	0

Table 3.8: Condition screen – Enzyme loading. Conditions taken forward are shown in yellow. Conditions: 2 mM CDG, 4 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 5 hours.

It was found that 2 mg/mL of BsQueC provides the best conversion to $PreQ_0$. Even at this concentration, the conversion to $PreQ_0$ was not total, although complete consumption of the starting material was observed. In addition, while even small enzyme concentrations lead to significant formation of the amide, ADG dehydration is more challenging, and ADG is, in all cases, the major product. It was also noted that, as expected, no reaction was observed when no enzyme was present.

3.5.1.2 MgATP equivalents

The second parameter which was investigated was the number of equivalents of MgATP. The conversion was measured by HPLC after 5 hours, as shown in Table 3.9. HPLC chromatograms after 5 hours reaction are shown on Figure 3.39.

Number of MgATP equivalents	MgATP concentration (mM)	CDG (%)	ADG (%)	$PreQ_0$ (%)
2	4	0	58	42
1.5	3	4	71	24
1	2	9	79	12
0.5	1	46	54	0
0	0	100	0	0

Table 3.9: Condition screen – MgATP equivalents. Conditions taken forward are shown in yellow. Conditions: 2 mg/mL BsQueC, 2 mM CDG, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 5 hours.

Winkler and co-workers suggested an adenylating mechanism where one equivalent of MgATP is required for the transformation of CDG to ADG and another MgATP equivalent to proceed to $PreQ_0$[9]. Therefore, two equivalents of MgATP would be needed to convert CDG into $PreQ_0$.

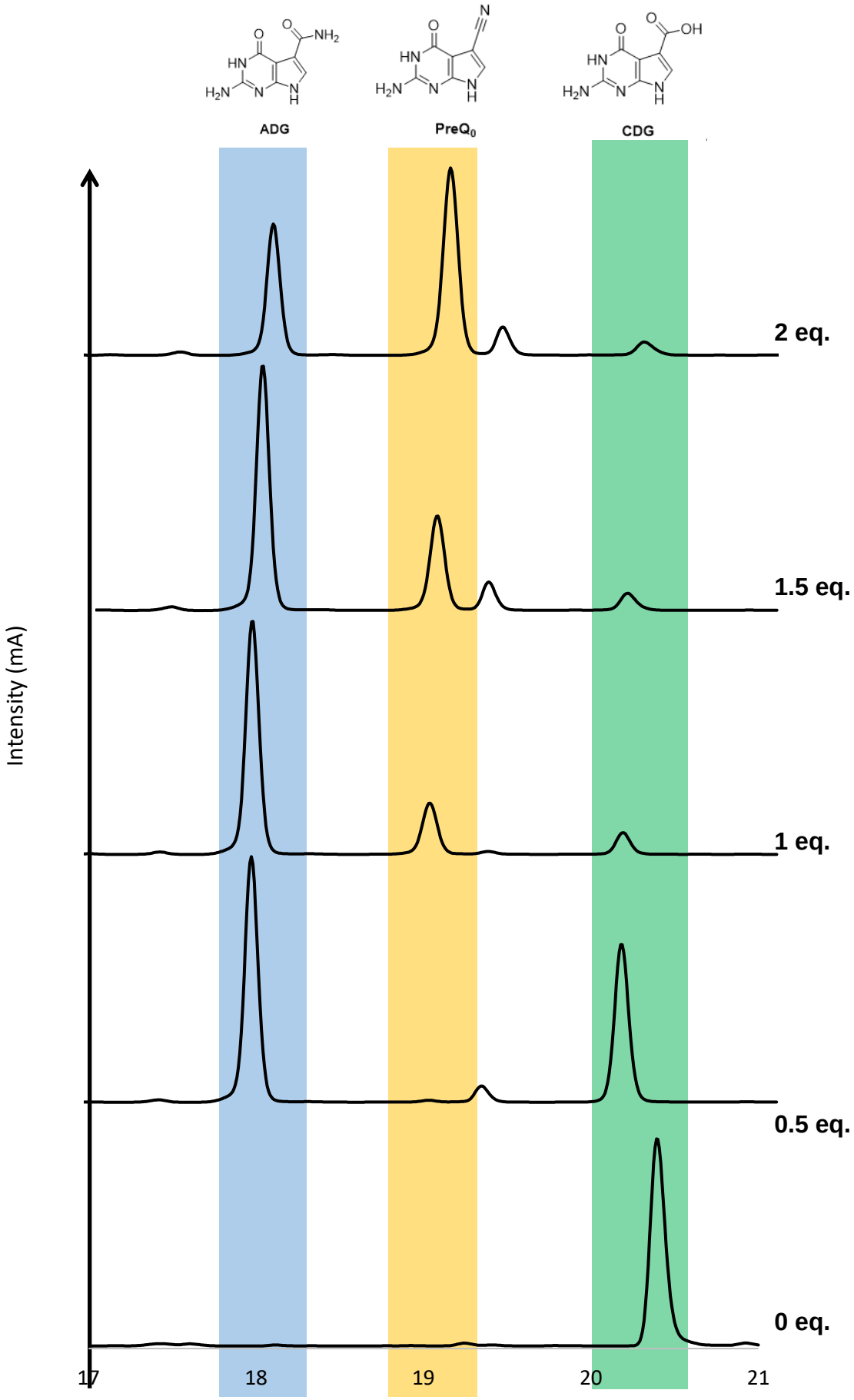


Figure 3.39: Reaction chromatograms after 5 hours for the NS reaction with multiple concentrations of ATP. Conditions: 2 mg/mL BsQueC, 2 mM CDG, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

As such, the results obtained seem fairly consistent with the mechanism postulated by Winkler *et al.*, where two equivalents of MgATP are needed to convert CDG into $PreQ_0$.

3.5.1.3 CDG loading

The substrate loading was the final parameter to be investigated. Different concentrations of CDG were investigated. Two equivalents of MgATP were added in each case. CDG conversion was assessed after 5 hours, as shown in Table 3.10.

CDG concentration (mM)	MgATP concentration (mM)	CDG (%)	ADG (%)	$PreQ_0$ (%)
10*	**20***	100	0	0
5*	**10***	100	0	0
2*	**4***	0	58	42
1	**2**	0	19	81
0.5	**1**	0	0	100

Table 3.10: Condition screen – CDG loading. Conditions taken forward are shown in yellow.
** Precipitation was observed upon MgATP addition.*
Conditions: 2 mg/mL BsQueC, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 5 hours.

Hence, the optimal CDG loading was found to be 0.5 mM. While at 4 mM MgATP, partial precipitation of the enzyme has been likely to affect the reaction, at higher concentrations of MgATP, full enzyme precipitation was observed. For this reason, no conversion of CDG was observed. A time-dependent study of the conversion of CDG to ADG and $PreQ_0$ using these optimised conditions is shown in Figure 3.40. The relative percentages of each species at all these time points are shown on Figure 3.41.

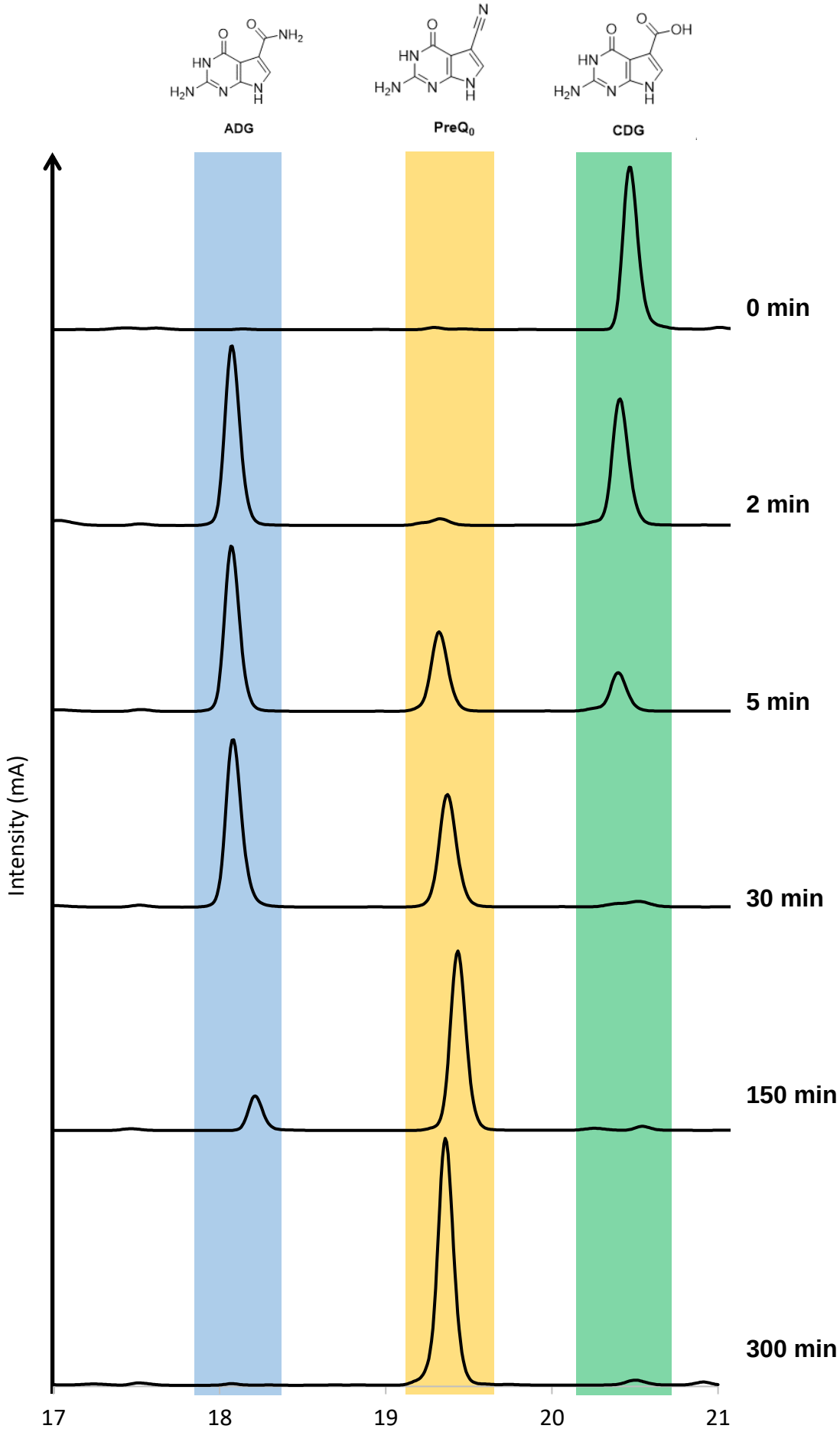


Figure 3.40: Time-dependant study of the NS-catalysed reaction.
Conditions: Conditions: 2 mg/mL BsQueC, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

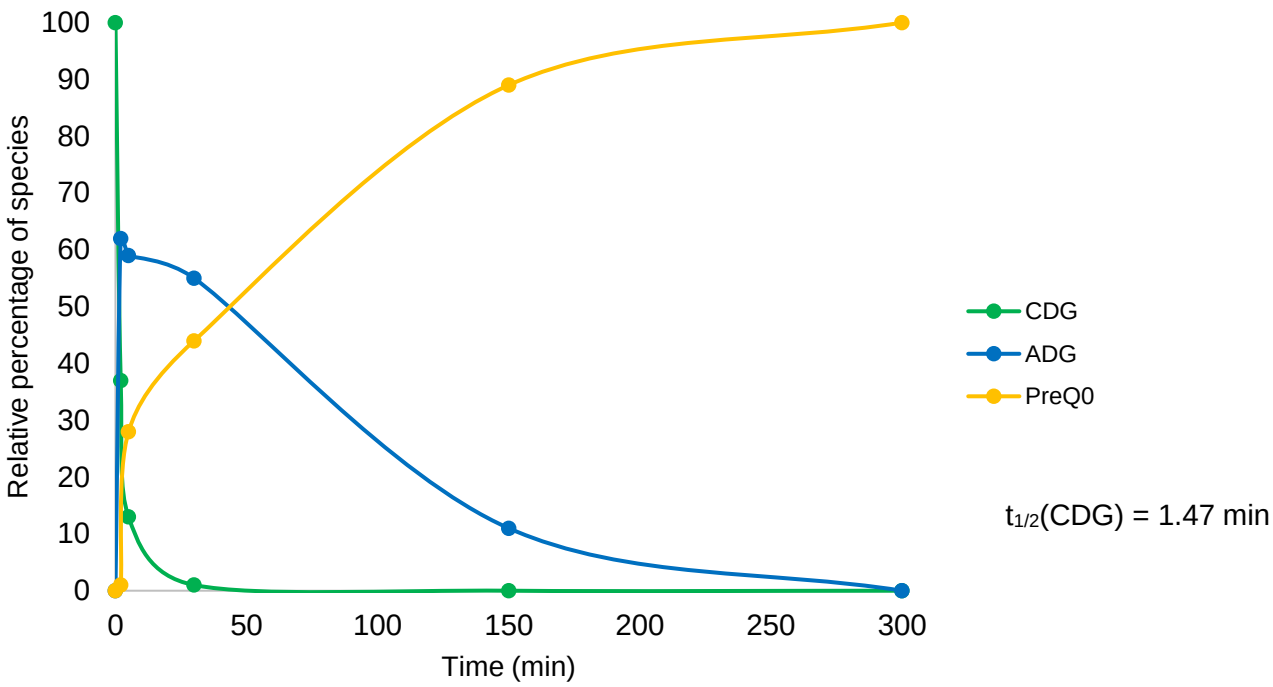


Figure 3.41: Evolution of the relative percentage of CDG, ADG and $PreQ_0$.
Conditions: Conditions: 2 mg/mL BsQueC, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM amine in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

As shown on Figures 3.40 and 3.41, the conversion of CDG to ADG is very quick, and the half-life of CDG was found to be 1.47 min and the full conversion of CDG to ADG is observed after 30 mins. In contrast, the conversion of ADG to $PreQ_0$ is much slower. This means that there is a very fast buildup of ADG that is converted into $PreQ_0$ over the full course of the reaction. The full disappearance of the ADG peak, that will be fully converted into $PreQ_0$ is observed after 300 mins.

SrToyM was also assayed using these optimised conditions. In this case, incomplete conversion of CDG was observed (97% conversion). ADG was the major product (63%) and $PreQ_0$ was minor (34%) as shown on Figure 3.42.

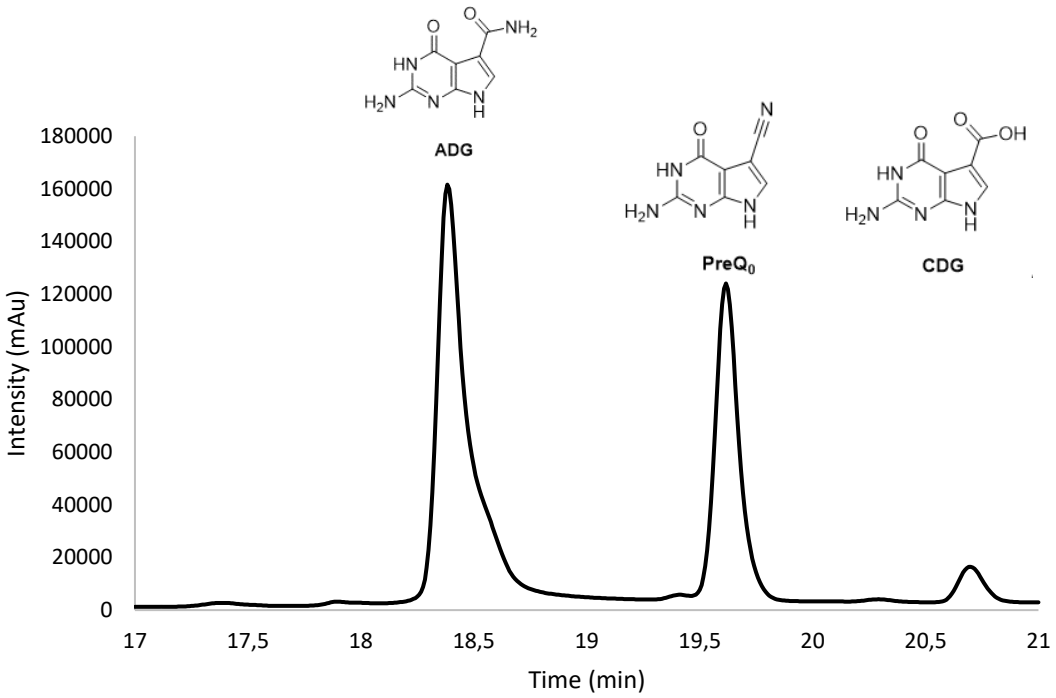


Figure 3.42: HPLC chromatogram of the SrToyM-catalysed synthesis of ADG and $PreQ_0$ from CDG after 5 hours. Conditions: 2 mg/mL SrToyM, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

3.6 Substrate screens

3.6.1 Carboxylic acid substrate screen

Winkler and colleagues have published a small screen of 11 carboxylic acids[1]. They had screened a few benzoic and heterocyclic acids to assess its scope which was discussed earlier in part 3.1.3. They concluded that BsQueC displays catalytic activity for no substrate other than CDG. No substrate scope studies have been carried out for SrToyM. A screen of other relevant carboxylic acid substrates with BsQueC and SrToyM is presented on Figure 3.43, and the results for compounds who displayed activity are shown on Table 3.11 (compounds that did not display activity are not shown on Table 3.11). HPLC chromatograms of the reactions after 5 hours for the substrates that displayed activity are shown on Figure 3.44.

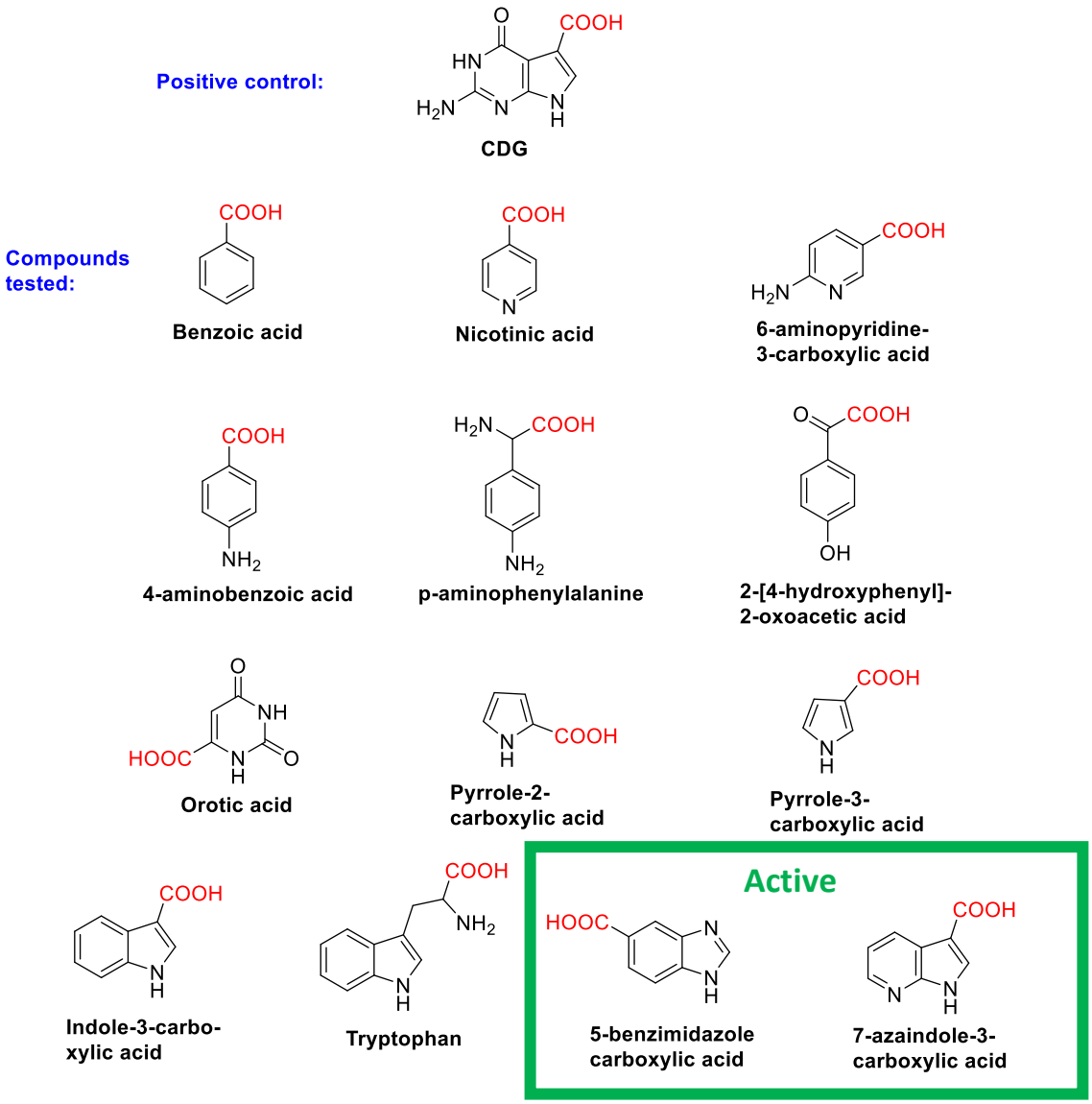


Figure 3.43: Carboxylic acid substrate scope studies of BsQueC and SrToyM. Conditions: 2 mg/mL BsQueC, 0.5 mM acid, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 24 hours.

	BsQueC conversion (%)	SrToyM conversion (%)
CDG (positive control)	100	100
5-benzimidazole carboxylic acid	94	8
7-azaindole-3-carboxylic acid	11	8

Table 3.11: Results for the compounds which displayed activity for the NSs tested. Compounds which did not display any activity are not shown.

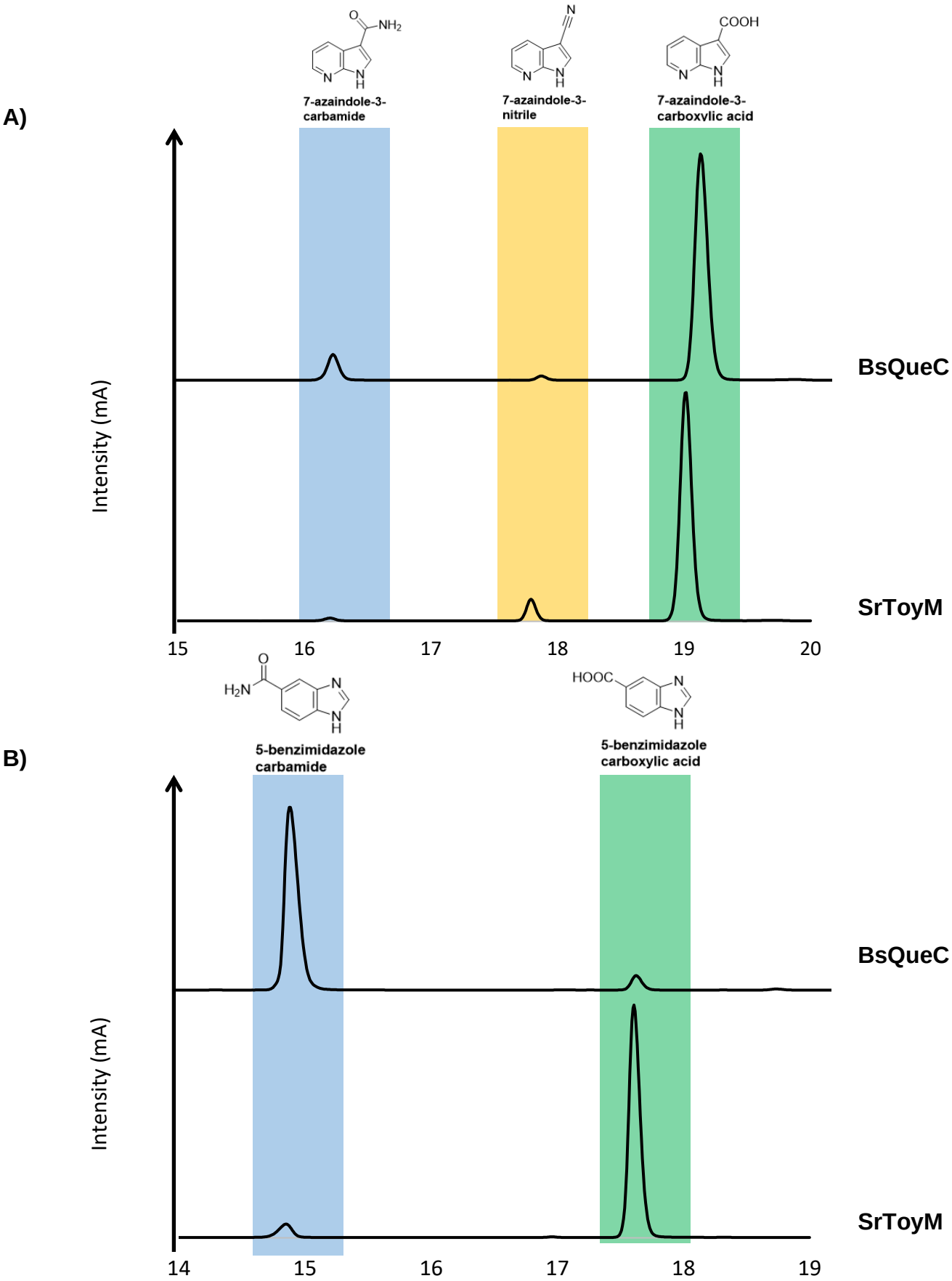


Figure 3.44: HPLC chromatogram for the NS-catalysed reactions of A) 7-azaindole-3-carboxylic acid and B) 5-benzimidazole carboxylic acid .
Conditions: Conditions: 2 mg/mL NS, 0.5 mM acid, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

Hence, this screen confirmed that the NSs carboxylic acid substrate scope is very restricted as all benzoic acids and most heterocyclic acids screened were not converted by either of the NSs. The only two compounds which were accepted by the NSs were 5-benzimidazole carboxylic acid and 7-azaindole-3-carboxylic acid. In common with CDG, both have an H-bond donor (N-H) at close proximity (2 bonds away) of an H-bond acceptor (N). As such, this might be a feature required for NS activity. This suggestion is strengthened by the fact that, in their substrate screen, Winkler showed that BsQueC does not exhibit any activity for 2-methyl-4-oxo-4,5,6,7-tetrahydroindole-3-carboxylic acid, shown on Figure 3.9, which is similar to CDG but does have one H-bond donor at close proximity of an H-bond acceptor[1]. The explanation of this pattern should be that these H-bonds donor and acceptor should be involved in critical interactions with the enzyme that potentially "lock" the compound in the active site to initiate the NS reaction. However, 6-aminopyridine-3-carboxylic acid, which was not converted by any NSs, also has a hydrogen bond donor (primary amine -NH_2) two bonds away from a hydrogen bond acceptor (N). Hence, suggesting that other features on the substrate are required for activity. More insights on the structure-activity relationship would be gained from a ligand-bound crystal structure as it would allow to precisely locate the residues interacting with the ligands.

3.6.2 Amine screen

It was shown in this work and also in the literature[1] that using ammonium sulfate in the NS-catalysed transformation of CDG generates ADG, which is readily converted by the NSs into the corresponding nitrile as ADG is a primary amide. However, using more complex amines would generate secondary or tertiary amides in lieu of ADG, which could not be transformed into the corresponding nitrile. In this part, some primary and secondary amines were screened to investigate the range of secondary and tertiary amides that could be generated using NS enzymes. This screen is presented in Figure 3.45, and the results for compounds who displayed activity are shown in Table 3.12 (compounds that did not display activity are not shown). The HPLC chromatograms of the SrToyM-catalysed reactions are shown on Figure 3.46.

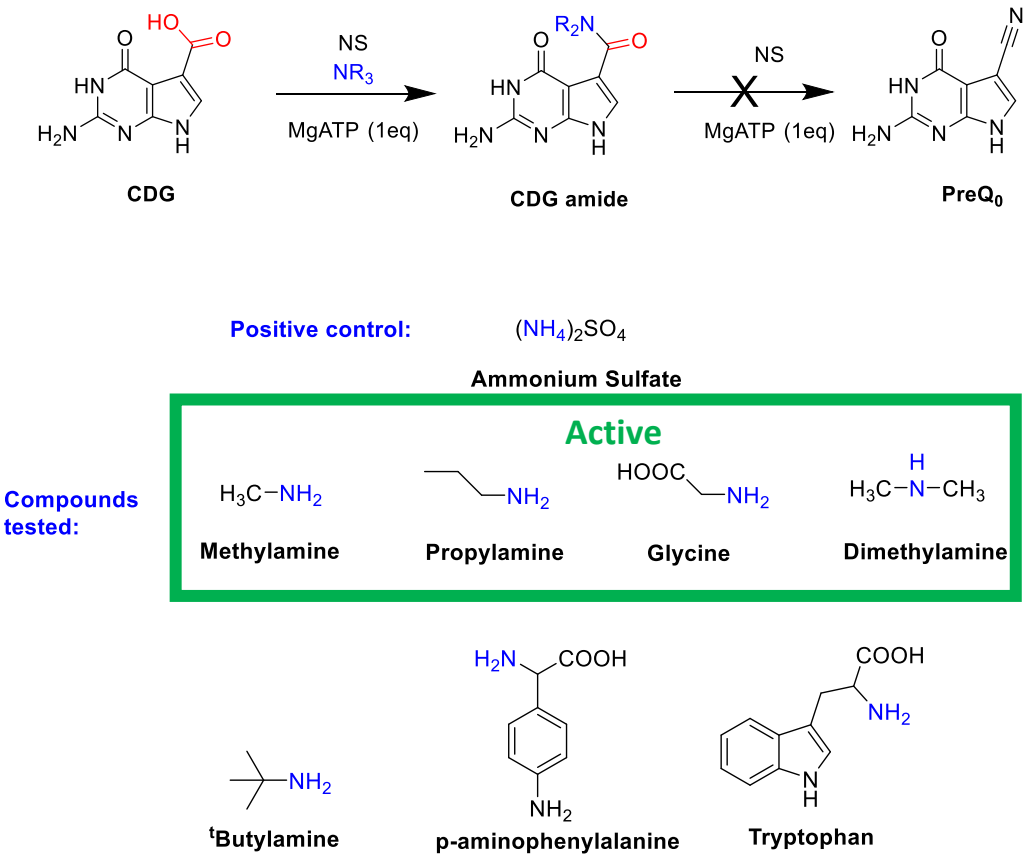


Figure 3.45: Amine scope studies of BsQueC and SrToyM.
Conditions: 2 mg/mL BsQueC, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM amine in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 24 hours.

	BsQueC conversion (%)	SrToyM conversion (%)
Ammonium sulfate (positive control)	100	100
Methylamine	81	68
Propylamine	79	85
Dimethylamine	70	72
Glycine	35	35

Table 3.12: Results for the compounds which displayed activity for the NSs tested. Compounds which did not display any activity are not shown.

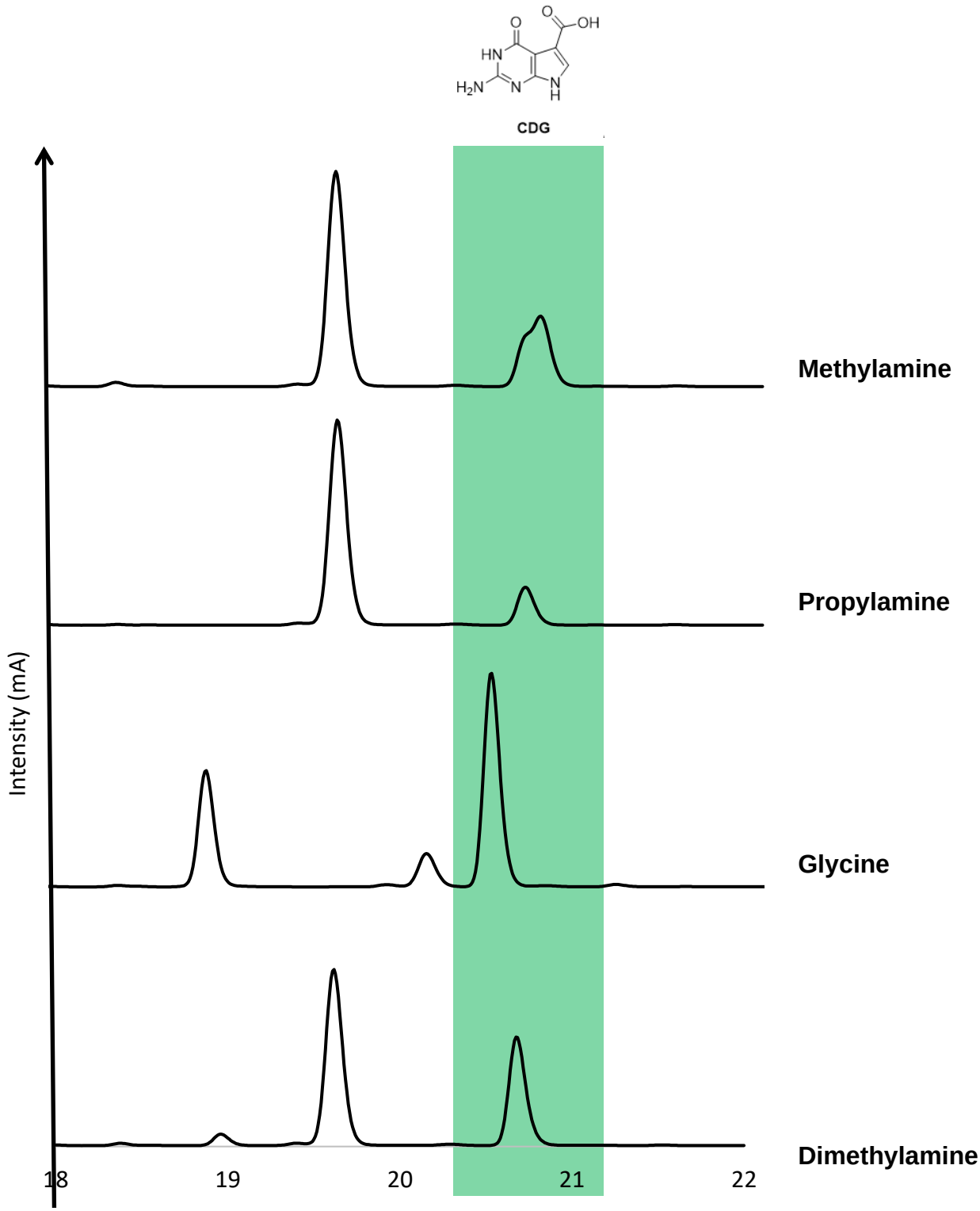


Figure 3.46: HPLC chromatogram for the SrToyM-catalysed reactions of CDG with various amines Conditions: Conditions: 2 mg/mL SrToyM, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM amine in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

Hence, it was observed that a small range of simple secondary and tertiary CDG-amides could be made using NSs enzymes as methylamine, dimethylamine, glycine and propylamine are accepted by NSs enzymes with conversions between 35-85%. However, none of the tested NSs exhibits activity towards bulkier amines, suggesting that the active site can only accommodate small amines. More insights on the structure-activity relationship would be gained from a ligand-bound crystal structure.

3.7 Structural studies

3.7.1 BsQueC

Structural studies of the NS enzymes were carried out to gain better mechanistic insights on this enzyme family. As discussed, two crystal structures of NS enzymes exist in the PDB, including one of BsQueC, determined by Cicmil and Huang[15]. However, on this structure, shown in Figure 3.47, zinc, magnesium and phosphate ions are bound to the enzyme, and neither CDG nor ATP is present.

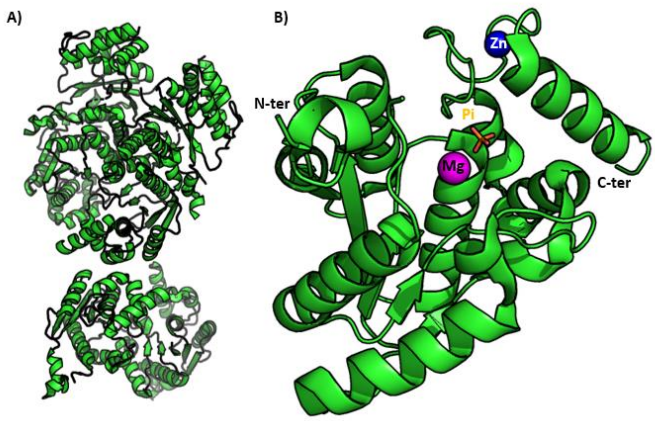


Figure 3.47: A) Structure of the BsQueC tetramer (PDB ID: 3BL5[15]).
B) Structure of a BsQueC monomer Zn^{2+} is shown as a blue sphere and $Mg^{2+}$is shown as a pink sphere. The phosphate ion is shown in sticks.

The aim was to obtain a structure of a NS with ligands to locate the active site.

3.7.1.1 HT conditions screens on the unbound enzyme

After unsuccessful replication of the conditions optimised by Cicmil *et al.*, HTP conditions screens were carried out to find suitable conditions to prepare diffraction-quality crystals. These screens were undergone using a Gryphon robot on a 96 wells plate with three different screens:

- **Molecular Dimensions JCSG+**: a broadly used screen for initial screening experiments as it contains a wide range of PEGs, salts, organic precipitants and neutralised carboxylic acids across a pH range from 3.0 to 10.5.

- **Molecular Dimensions Structures 1+2:** a screen developed by Jancarik and Kim where the composition of each crystallisation solutions was empirically derived from common previously published conditions[53].
- **Hampton Research PEG/Ion:** a screen where each crystallisation solution is composed of polyethylene glycol (PEG) 3350 and salts representing a complete range of anions and cations commonly used for the crystallisation of proteins.

Crystals were generated by vapour diffusion *via* the sitting drop method. Initial screens, with an enzyme concentration of 15 mg/mL, were unsuccessful as the formation of protein aggregates was observed. However, at an enzyme concentration of 7.5 mg/mL, one encouraging hit, from the Molecular Dimensions JCSG+ screen was observed, as shown in Figure 3.48.

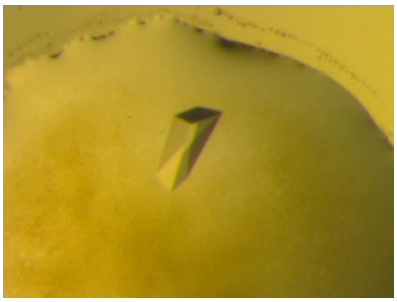


Figure 3.48: HT condition screens hit for BsQueC.
Conditions: BsQueC: 7.5 mg/mL; 0.2 M magnesium chloride hexahydrate, 0.1 M Tris (pH 7), 10% w/v PEG 8000.

3.7.1.2 Optimisation

These hit conditions were then optimised both for ligand-free BsQueC, (BsQueC + MgADP), (BsQueC + MgATP), (BsQueC + CDG) and (BsQueC + CDG + MgATP). The crystals were generated by vapour diffusion *via* the hanging drop method and grown at 23°C. Optimisation conditions are shown on Figure 3.49 and crystals obtained after 7 days growth are shown in Table 3.13.

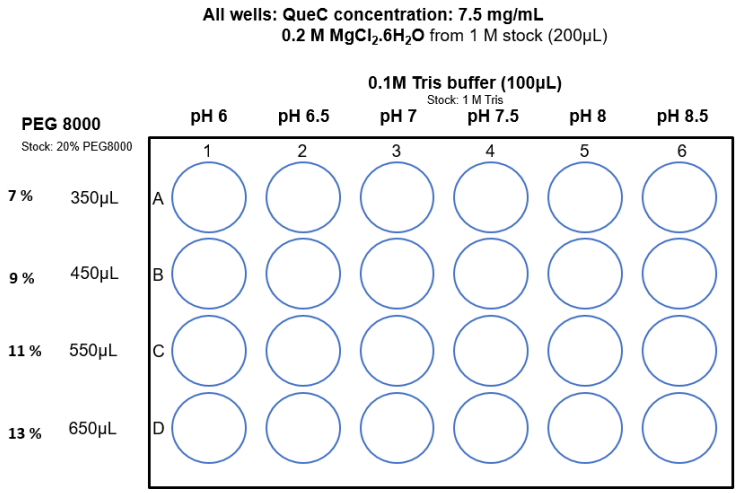
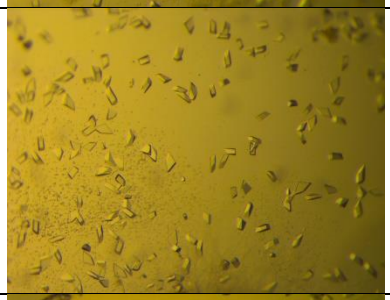
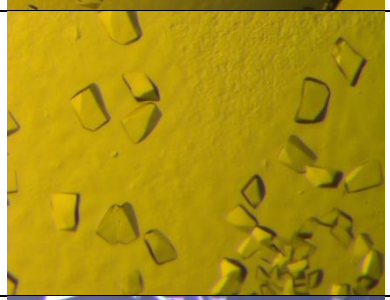
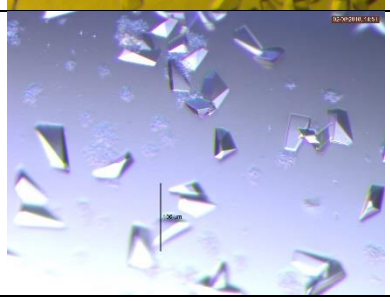


Figure 3.49: Optimisation conditions for BsQueC.

Experiment	Additives	No	Conditions	Size	Crystal morphology	Outcome
BsQueC ligand-free	None	1	A1	<40 Å		Too small
		2	A4	40-60 Å		Diffraction to >4 Å
		3	B5	40-60 Å		Diffraction to >4 Å
BsQueC + MgADP	2.5 mM ADP 12.5 mM $MgCl_2$	4	A3	80-100 Å		Diffraction to 2.1 Å
		5	A4	60-80 Å		Diffraction to >4 Å
		6	A3	80-100 Å		No diffraction

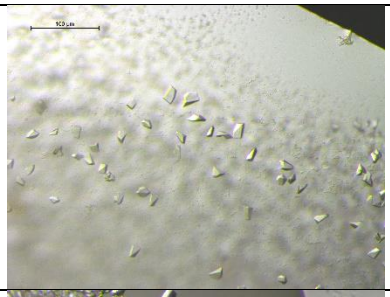
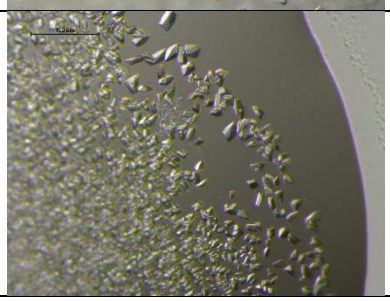
Experiment	Additives	No	Conditions	Size	Crystal morphology	Outcome
BsQueC + MgATP	2.5 mM ATP 12.5 mM $MgCl_2$	7	C3	20-40 Å		Too small
		8	C1	20-40 Å		Too small
BsQueC + CDG	1.25 mM CDG (in DMSO)	9	N/A	N/A	N/A	No crystals formed
	0.5 mM CDG (in DMSO)	10	N/A	N/A	N/A	No crystals formed
	0.5 mM CDG (in 500 mM KOH)	11	N/A	N/A	N/A	No crystals formed
	CDG (powder)	12	N/A	N/A	N/A	No crystals formed
BsQueC + CDG + MgATP	1.25 mM CDG (in DMSO) 2.5 mM ATP 12.5 mM $MgCl_2$	13	N/A	N/A	N/A	No crystals formed
	0.5 mM CDG (in DMSO) 2.5 mM ATP 12.5 mM $MgCl_2$	14	N/A	N/A	N/A	No crystals formed

Table 3.13: Results of BsQueC crystallisation optimisation.

Hence, crystals of ligand-free BsQueC, BsQueC + MgADP and BsQueC + MgATP were obtained. Most crystals were of rhombohedron shape with also some tetrahedron- or scalenohedron-shaped crystals. However, no crystals were obtained when CDG was present. Lowering CDG loading, changing CDG stock solvent to a strong base (500 mM KOH), where CDG is also soluble, or adding CDG powder to the drop was also unsuccessful. In addition, further HTP screens with BsQueC + 0.5 mM CDG in DMSO and in 500 mM KOH were undergone, but none resulted in any hits.

The crystals obtained were fished, prepared and diffracted in collaboration with Dr Jon Marles-Wright (Newcastle University). While most crystals did not exhibit any diffraction or showed poor resolution (> 4 Å), the crystals from Entry 4 displayed an adequate resolution of 2.1 Å and were used to solve the structure of BsQueC + MgADP.

3.7.1.3 BsQueC + ADP structure

Solving of the BsQueC + ADP structure was performed by Dr Jon Marles-Wright (Newcastle University) by molecular replacement using the BsQueC structure with PDB code 3BL5, shown in Figure 3.44. The structure reported in 3BL5 has a resolution slightly better than the one obtained in this study (2.95 Å for 3BL5 vs 2.1 Å here). There are six protein molecules in one asymmetric unit, four of which forming a tetramer while the two remaining form a tetramer with two other protein molecules from a neighbouring asymmetric unit. A monomer is shown in Figure 3.47. As expected, similarly to 3BL5 structure, a Rossmann fold composed of 5 β-sheets flanked by five α-helices can be observed between residues 3 and 155 as well as a phosphate-binding loop, between residues 9 and 14. In addition, a C-terminal zinc-binding motif is also featured with a cluster of 4 cysteine residues which should complex a Zn^{2+} ion. Interestingly, a disulphide bond between Cys186 and Cys195 was rendered. However, the electron density corresponding to that bond could also be the Zn^{2+} ion. Both ADP and a Zn^{2+} ion are thought to be bound to this structure as areas of electron density are observed both at the proximity of the phosphate-binding loop and the cysteine cluster as shown on Figure 3.50. Similarly to the 3BL5, structure determination was incomplete as the loop between residues 76-91 was not rendered, suggesting that this loop is very flexible. In addition, residues 123-128 were also part of a loop that was not rendered.

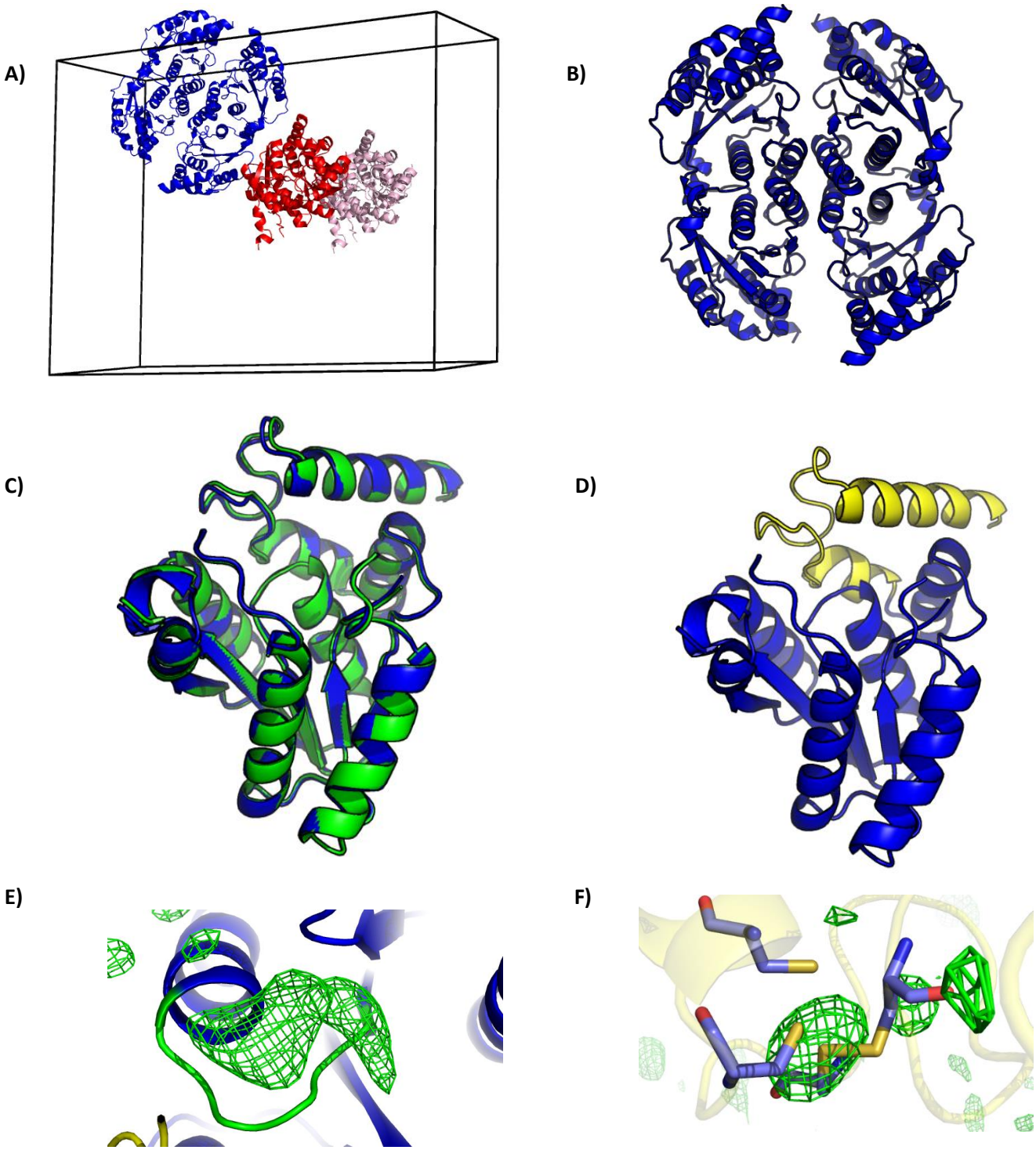


Figure 3.50: A) Structure of a unit cell in the final refinement of BsQueC structure. Six BsQueC molecules are part of each unit cells: four of them (blue) form a tetramer while the two other (red) form a tetramer with two BsQueC molecules (light pink) that are part of another unit cell
B) Structure of a BsQueC tetramer
C) Structure of a BsQueC monomer (blue) superposed with a BsQueC monomer from PDB structure 3BL5 (green)
D) Structure of a BsQueC monomer displaying the Rossmann fold (blue) and the Zn^{2+} binding site (yellow)
E) Area of electron density (green) thought to be MgADP bound to the phosphate-binding loop (light green)
F) Area of electron density (green) thought to be a Zn^{2+} ion in close proximity of the cysteine residues cluster

Coot, one of the programmes of the Collaborative Computational Project No. 4 (CCP4) suite was then used to fit an ADP ligand onto the electron density area shown in Figure 3.50-E. The ADP 3D structure was taken from the PDB database. Because of the fairly low resolution (3.1 Å), it is not fully certain that this electron density area corresponds to a ligand, as it could also correspond to another conformation of the phosphate-binding loop. However, it was found that ADP could fit the electron density area fairly well, as shown in Figure 3.51.

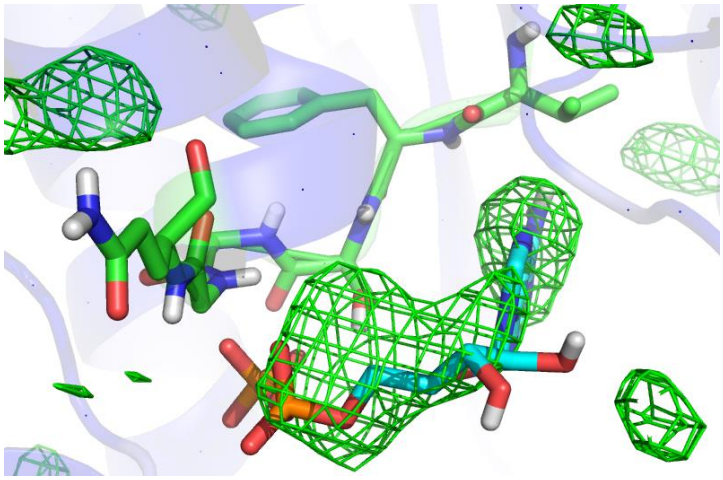


Figure 3.51: Structure of the ADP pose that is most closely matching the electron density area (cyan). The phosphate-binding loop residues (Val9-Gln14) are shown in green sticks.

Hence, as expected, ADP phosphate moiety faces the phosphate-binding loop (Val9-Gln14) shown in green sticks in Figure 3.51). These interactions are shown in Figure 3.52.

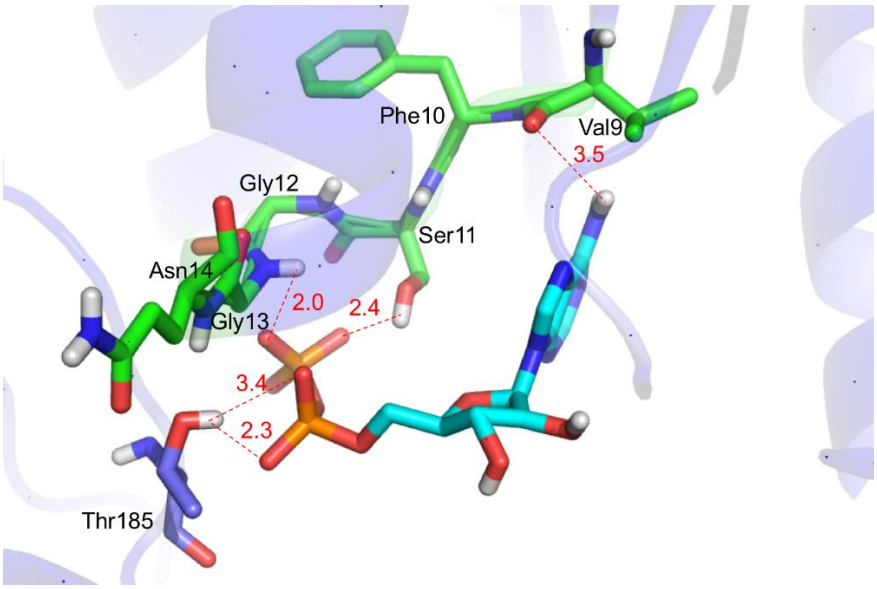


Figure 3.52: Interactions between ADP and BsQueC. The ADP molecule is shown in cyan sticks. The residues of the phosphate-binding loop (green) and also T185 (light blue) are also shown in sticks. The intermolecular distance in angstroms between H atoms of H-bond donors and the H-bond acceptors is shown in red.

Specifically, numerous H-bound interactions can be observed as the oxygen atoms which are part of the terminal phosphate group are involved in possible H-bond interactions with the Ser11 side chain hydroxyl group, Gly13 backbone NH proton as well as the Thr185 side chain hydroxyl, which is not part of the phosphate-binding loop. The Thr185 hydroxyl could also interact with the ADP α-phosphate group. Finally, the primary amine of the adenine moiety could also interact with the carboxyl oxygen atom of Val9, in the phosphate-binding loop.

The structure of the BsQueC:ADP-bound complex was compared with the BsQueC:phosphate complex previously determined by Cicmil *et al.* (PDB: 3BL5, Figure 3.47). Both the phosphate and the ADP are in close proximity to the phosphate-binding loop. In addition, it was found that the phosphate ion from the BsQueC:phosphate complex overlaps with the α-phosphate of ADP from the structure of the BsQueC;ADP complex. This is supporting evidence that the electron density that was used to fit the ligand corresponds to an ADP molecule.

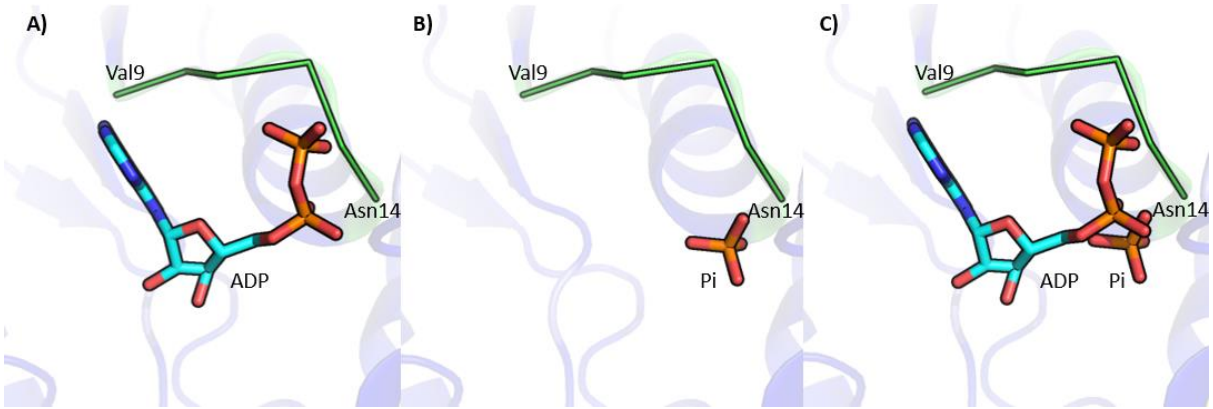


Figure 3.53: A) Structure of the BsQueC:ADP complex (determined in this work). B) Structure of the BsQueC:phosphate complex (PDB: 3BL5; determined by Cicmil et al.). C) Overlay of the two complexes using the Align feature of Pymol. The BsQueC has been removed leaving the phosphate ligand. The phosphate-binding loop is composed of residues Val9-Gln14 (green). Both the phosphate from 3BL5 structure and the ADP phosphate groups are in close proximity to the phosphate binding loop and the phosphate from 3BL5 overlaps closely with ADP α-phosphate group.

Looking back at the whole protein-ligand complex, the coulombic electrostatic surface map, shown on Figure 3.54 displays BsQueC binding pocket. The large active site cavity is fit to accommodate both MgATP, the substrate and the amine and a predominance of positively-charged residues is observed around the ADP molecule. This is not surprising as these positively-charged residues should be able to stabilise the negatively charged adenylate intermediates.

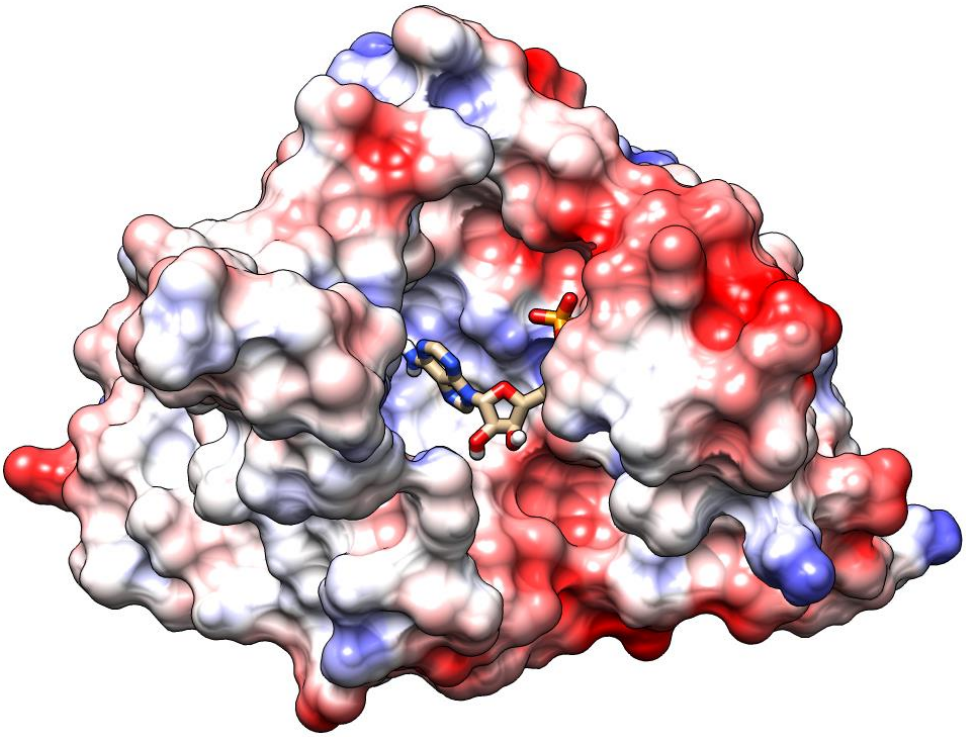


Figure 3.54:BsQueC surface coloured by electrostatic potential showing a large binding pocket which accommodates the ADP molecule.

3.7.2 SrToyM

3.7.2.1 HT conditions screens on the ligand-free enzyme

Unlike BsQueC, the crystallographic structure of SrToyM has not been determined. As such, high-throughput condition screens were carried out. The same screens than for BsQueC (Molecular Dimensions JSCG+ and Structures 1+2 and Hampton Research PEG/Ion) were chosen. These screens were carried with 10 mg/mL SrToyM. In this case, three hits, shown in Figure 3.55, were optimised.

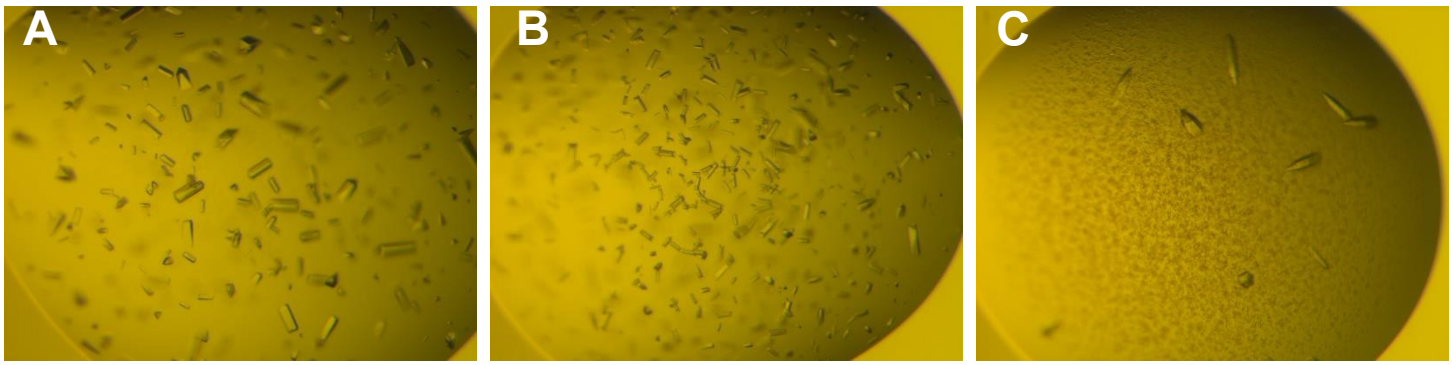


Figure 3.55: HT condition screens hits for SrToyM (10 mg/mL).
Conditions A: 1.26 M Ammonium sulfate, 0.2 M lithium sulfate, 0.1 M Tris pH 8.5
Conditions B: 1.6 M Ammonium sulfate, 10% w/v dioxane, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5
Conditions C: 20% w/v PEG 3350, 0.2 M sodium malonate pH 5

3.7.2.2 Generation of Selenomethionine-enriched SrToyM

As, unlike for BsQueC, no structure of SrToyM is available, solving the SrToyM structure might be challenging because of the phase problem. The data generated by an X-ray diffractometer is a diffraction pattern, which will show how the protein crystal diffracted the incident light beam[54]. In order to reconstruct the structure, the amplitude and the phases of the diffracted light waves need to be determined. However, only the amplitude of the light waves can be recovered from the measured intensity of the spots. For proteins that have high identity to previously solved proteins, molecular replacement (MR) allows easier determination of the phases through "using" the phases of the solved structure. Sequence analysis of BsQueC and SrToyM showed they display 30.7% identity, which should be high enough to carry MR to solve SrToyM structure, as MR can be successful with proteins that have an identity as low as 20%[54]. However, selenomethionine-enriched ToyM (SeMet-SrToyM) crystals were also generated. The incorporation of heavy atoms such as selenium allows solving the phase problem when MR is not possible. The diffraction maps originating from the WT and SeMet enriched proteins will be compared as all the differences between these two datasets will be caused by the heavy atoms. Their position in the protein will be determined and will be used as a starting point to determine the protein phase angles. As such, SeMet-SrToyM was expressed in B834(DE3) cells, a parental strain for BL21(DE3) which is methionine auxotroph and selenomethionine was added to the media. The enzyme was purified as described in Part 3.2.1 and characterised by SDS-PAGE and ESI-MS (Figure 3.56).

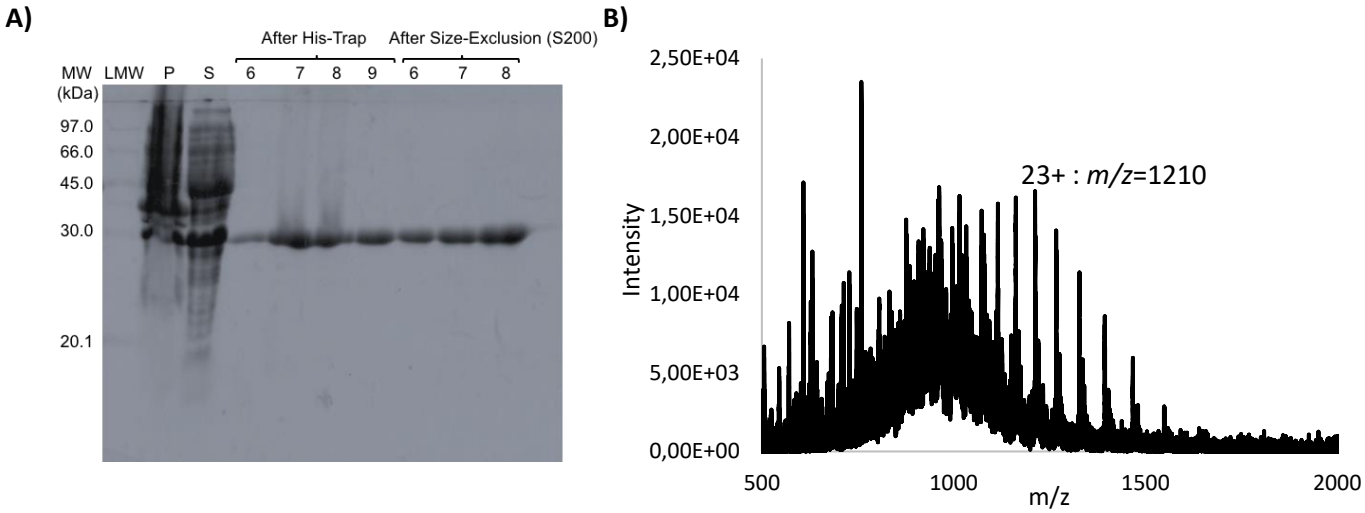


Figure 3.56: A) SDS-PAGE (P: pellet, S: soluble)
B) ESI-MS spectrum of SeMet-SrToyM. Experimental mass: 27801.10 ± 6.1Da

This ESI-MS analysis was useful to verify that SeMet residues had been fully incorporated in lieu of methionines. It has been previously determined that the molecular weight of SrToyM is 27498.34

± 1.12 Da and the molecular weight of the SeMet-SrToyM was found to be 27801.10 ± 6.19 Da. The molecular weight difference between SrToyM and SeMet-ToyM is 302.76 Da. This mass difference is due solely to the substitution of the sulfur atoms of methionine by selenium atoms in selenomethionine. With the atomic mass difference between selenium and sulfur being 46.9 Da, the number of incorporated selenomethionines was found to be 6. As there are five methionines in SrToyM sequence as well as one methionine in the part of the pET28a vector that is expressed, it was concluded that SeMet had been successfully incorporated at all the Met positions.

3.7.2.3 Optimisation of Conditions A

As for BsQueC, the hits that resulted from the HT screen were optimised. As conditions A gave the most symmetrical and largest crystals, these conditions were explored first and in most details for ligand-free SrToyM, ligand-free SeMet-SrToyM, (SrToyM + MgADP), (SrToyM + MgATP), (SrToyM + CDG) and (SrToyM + CDG + MgATP).). The crystals were generated by vapour diffusion *via* the hanging drop method and grown at 23C. The conditions are shown in Figure 3.57, and the crystals obtained after seven days are shown in Table 3.14.

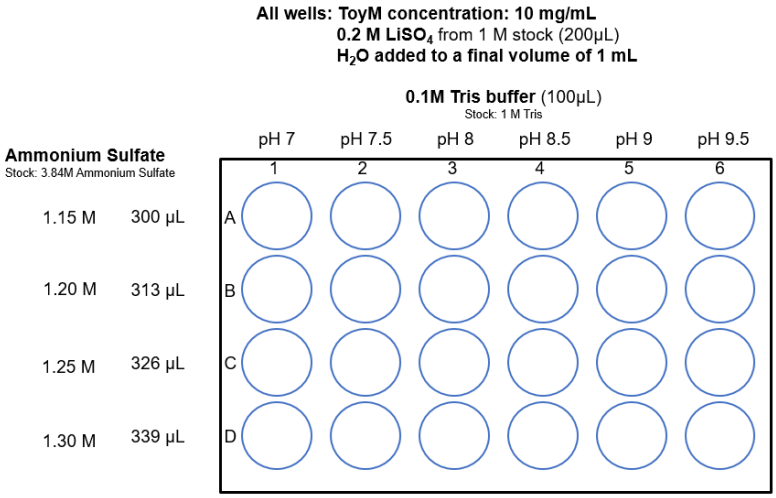
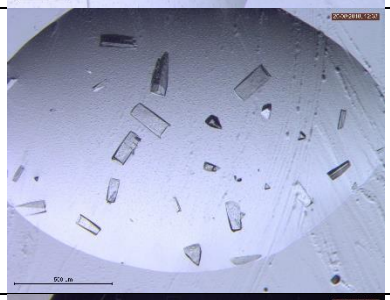
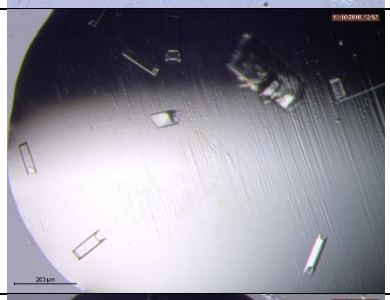
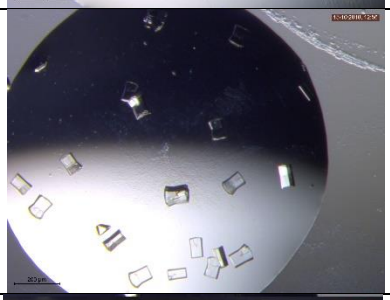
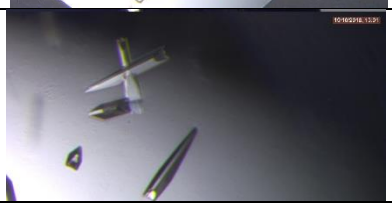


Figure 3.57: Optimisation conditions for Conditions A.

Experiment	Additives	No	Conditions	Size	Picture	Outcome
SrToyM ligand-free	None	1	A2	100-150 Å		No diffraction
		2	B1	150-200 Å		Diffraction to >4 Å
		3	C1	150-200 Å		Diffraction to >4 Å

Experiment	Additives	No	Conditions	Size	Picture	Outcome
SeMet-SrToyM ligand-free	None	4	A1	200+ Å		No diffraction
		5	B2	200+ Å		No diffraction
		6	D1	150-200 Å		No diffraction
SrToyM + MgADP	2.5 mM ADP 12.5 mM $MgCl_2$	7	A1	100-150 Å		No diffraction
		8	A2	100-150 Å		No diffraction
		9	A3	100-120 Å		No diffraction

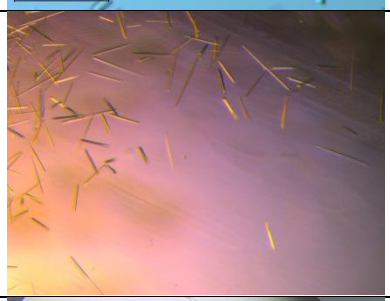
Experiment	Additives	No	Conditions	Size	Picture	Outcome
SrToyM + MgATP	2.5 mM ATP 12.5 mM $MgCl_2$	10	A1	100-150 Å		No diffraction
		11	A3	50-100 Å		Diffraction to >4 Å
		12	C2	150-200 Å		Diffraction to >4 Å
		13	C3	150-200 Å		No diffraction
SrToyM + CDG	0.5 mM CDG (in DMSO)	14	N/A	N/A	N/A	No crystals formed
SrToyM + CDG + MgATP	0.5 mM CDG (in DMSO) 2.5 mM ATP 12.5 mM $MgCl_2$	15	N/A	N/A	N/A	No crystals formed

Table 3.14: Results of SrToyM and SeMet-SrToyM crystallisation optimisation with Conditions A.

The crystals formed with these conditions have a prismoid shape. In the ligand-free enzyme trials (both SrToyM and SeMet-SrToyM) the crystals appear to be irregular at both ends. However, this was not observed upon addition of MgADP or MgATP. Interestingly, some SrToyM-MgATP co-crystals (conditions A1 and C2) were much thinner than the other crystals, adopting a needle-shape. This could be a sign that that MgATP had successfully bound the active site of the enzyme. However, this could

not be verified because none of the crystals tested diffracted to an adequate resolution. As for BsQueC, no crystals were obtained for when CDG was added. Other conditions found in the HTP condition screen described in Part 3.7.2.1 were then optimised.

3.7.2.4 Optimisation of Conditions B

Conditions B were also optimised, as shown in Figure 3.58.

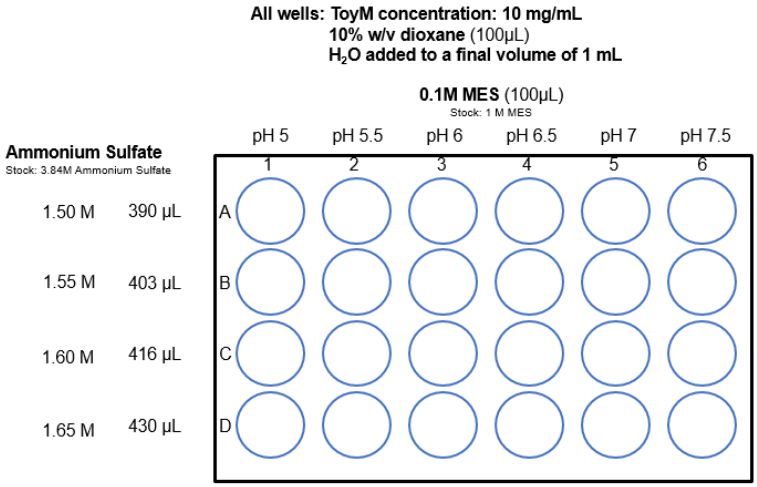


Figure 3.58: Optimisation conditions for Conditions B.

These conditions were tried both to obtain crystals of ligand-free SrToyM and co-crystals of SrToyM + MgADP and SrToyM + MgADP. However, after seven days of growth at 23°C, no crystals were observed in any of the conditions.

3.7.2.5 Optimisation of Conditions C

Finally, Conditions C obtained from the condition screen were optimised for ligand-free SrToyM, (SrToyM + MgATP) and (SrToyM + CDG + MgATP) as on Figure 3.59. Results are shown in Table 3.15.

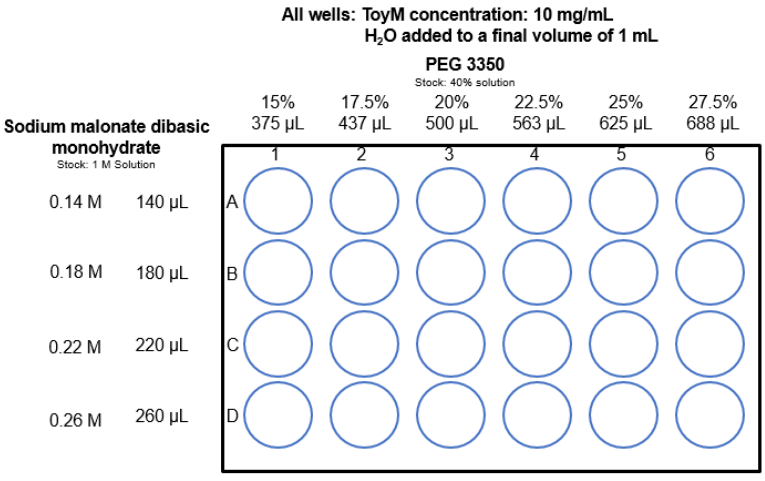


Figure 3.59: Optimisation conditions for Conditions C.

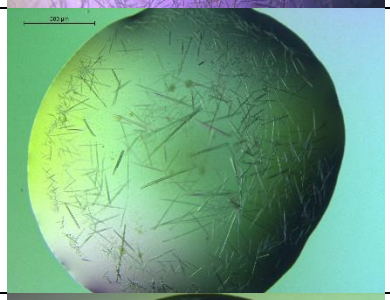
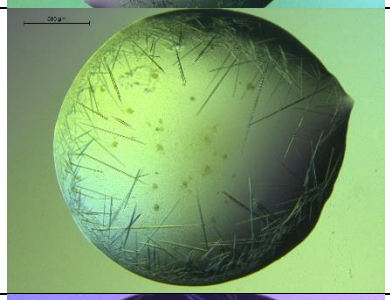
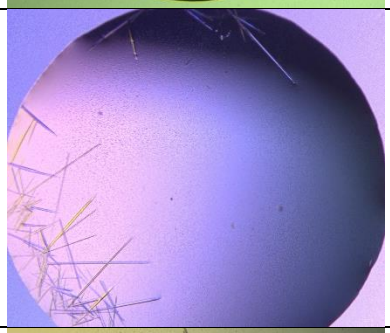
Experiment	Additives	No	Conditions	Size	Picture	Outcome
SrToyM ligand-free	None	1	A2	100-150		No diffraction
SrToyM + ATP	2.5 mM ATP 12.5 mM $MgCl_2$	2	A1	500-600		No diffraction
		3	B2	500-600		No diffraction
		4	B3	600-700		No diffraction
		5	D5	300-400		No diffraction
SrToyM + CDG	0.5 mM CDG (in DMSO)	6	N/A	N/A	N/A	No crystals formed
SrToyM + CDG + MgATP	0.5 mM CDG (in DMSO) 2.5 mM ATP 12.5 mM $MgCl_2$	7	N/A	N/A	N/A	No crystals formed

Table 3.15: Results of SrToyM crystallisation optimisation with Conditions C

It was noted that, while very large needle-shaped crystals were obtained for both unbound SrToyM and SrToyM + MgATP, none of the obtained crystals diffracted. In addition, as before, no SrToyM co-crystals were obtained upon adding CDG.

3.7.2.6 SrToyM + CDG co-crystals

Obtaining SrToyM + CDG co-crystals was unsuccessful so far. For this reason, HT conditions screens were done for SrToyM + 1 mM CDG + 2 mM MgATP + 10 mM magnesium chloride. The same screens than the ones described in 3.5.1.1 were used (Molecular Dimensions JSCG+ and Structures 1+2 and Hampton Research PEG/Ion). One hit was obtained, as shown in Figure 3.60.

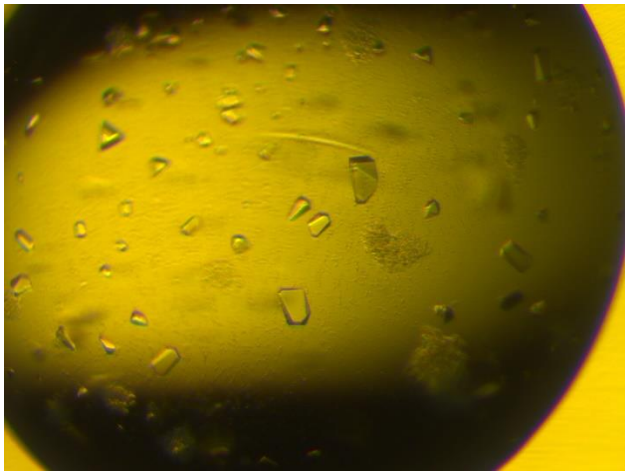


Figure 3.60: HT condition screens hit for SrToyM + CDG + MgATP + $MgCl_2$.
Conditions: SrToyM: 10 mg/mL., 1 mM CDG, 2 mM MgATP, 10 mM magnesium chloride; Precipitant: 0.1 M bicine (pH 9), 20% w/v PEG 6000.

These initial hit conditions were optimised, as shown in Figure 3.61:

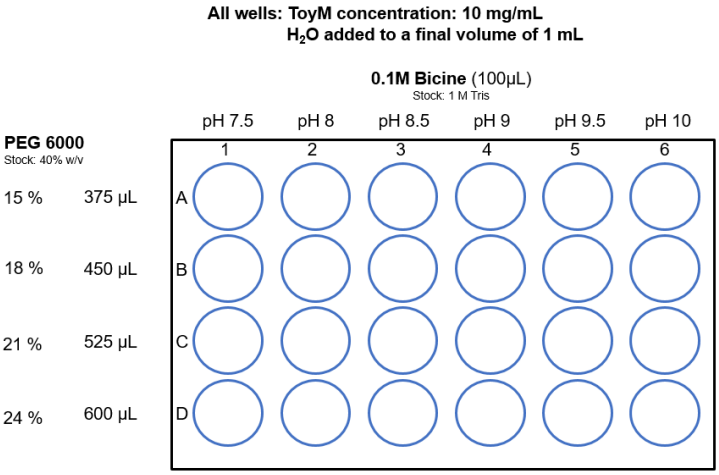


Figure 3.61: Optimisation conditions for the hit obtained after HT conditions screening for ToyM + CDG + MgATP + $MgCl_2$.

Optimisation of these conditions failed to yield crystals. Other attempts through reducing CDG concentration to 0.1 mM, changing CDG stock solvent to 500 mM KOH or not adding ATP and $MgCl_2$ were also unsuccessful. This suggests that the crystals obtained in the HT screen might have been precipitant salts.

3.7.2.7 Structure of SrToyM

Because none of the crystals obtained diffracted to a good resolution, the structure of SrToyM was predicted using RaptorX, using the BsQueC structure published in the PDB (3BL5) as template[15]. This program predicts secondary and tertiary structures of proteins from a previously-determined structure of high sequence similarity[55]. The predicted structure of SrToyM is shown in Figure 3.62.

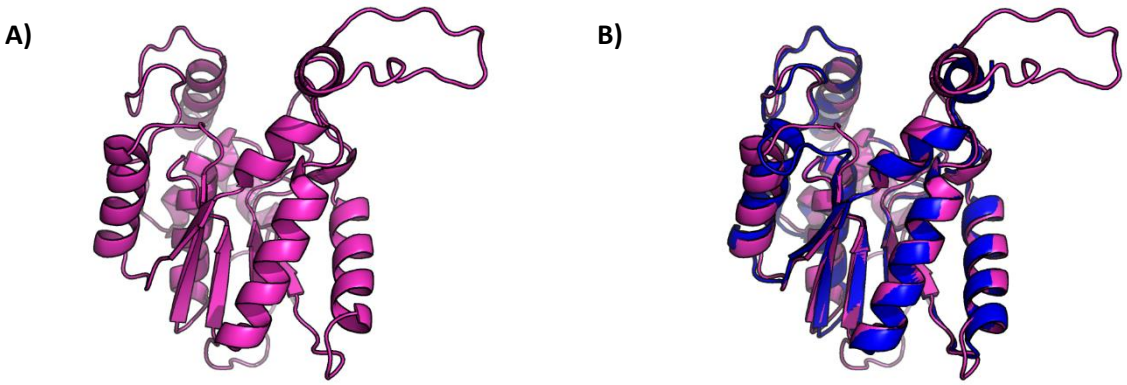


Figure 3.62: A) RaptorX-predicted structure of SrToyM
B) RaptorX-predicted structure of SrToyM (magenta) superposed with BsQueC structure (blue)

Hence, as shown in Figure 3.62 B), the predicted structure of SrToyM is very similar to the one of BsQueC. SrToyM is predicted to also feature a Rossmann fold as well as a Zn binding domain. The loop that was rendered neither on 3BL5 nor on the BsQueC structure presented in this work was predicted here (residues 86 to 98 of SrToyM) and can be seen on the top right of Figure 3.62.

3.8 Engineering an amide synthetase

So far, the exact location of the active site for the NS-catalysed reactions is unknown, and no mechanistic studies of the reactions have identified catalytically essential residues. As such, locating the active site is key for NS biocatalytic engineering. The location of the MgATP binding site could be inferred from the published structure as both magnesium and phosphate are bound to it and was then further confirmed in this project as the first ADP-bound structure of BsQueC was solved as shown on Figures 3.50-3.52. In order to choose the target residues for enzyme engineering, residues that fulfil the following three conditions were searched on BsQueC:

- **Residues with hydrophilic side chains, well conserved across all species**
- **Not involved in the binding of MgATP or Zn^{2+}**
- **Close in space to the ADP pyrophosphate moiety binding site**

3.8.1 Identifying conserved residues

To identify residues potentially involved in substrate binding and catalysis, sequence homology was performed to highlight conserved residues. This search was done as residues that are involved in substrate binding, and catalysis should be well conserved. Besides, considering CDG and ADG high number of H-bond donors and acceptors, it was postulated that hydrophilic residues should strongly favour substrate binding. As such, the sequence alignment of some NSs, showing all well conserved hydrophilic residues is shown in Figure 3.63.

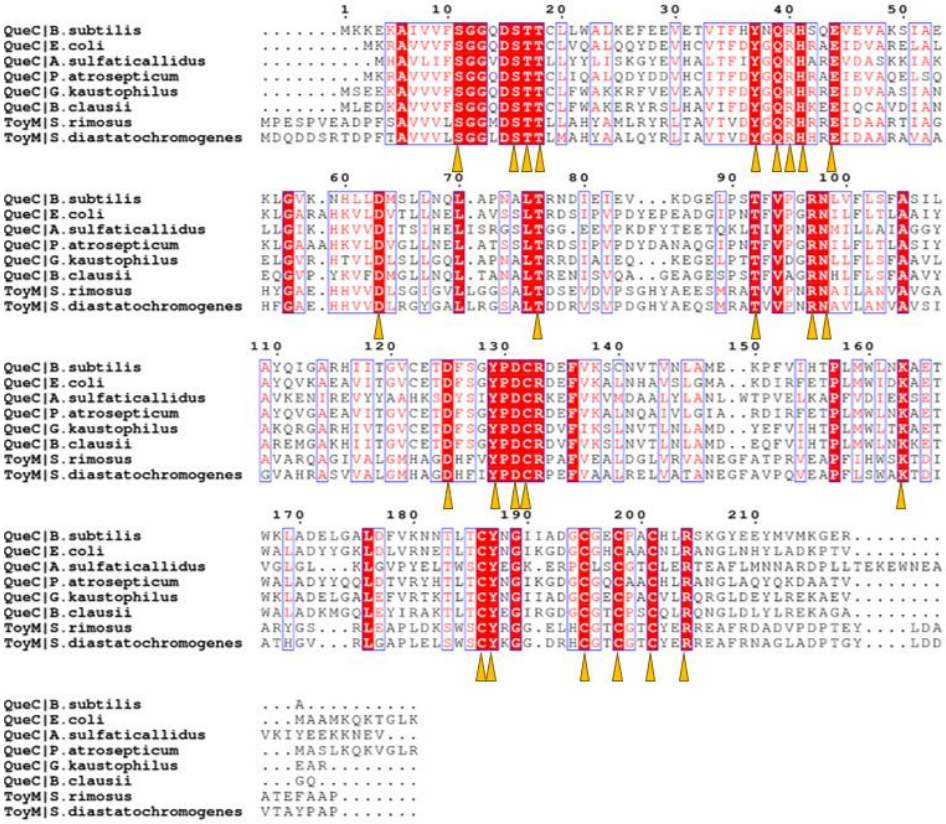


Figure 3.63: Sequence alignment the NSs from B. subtilis (BsQueC) (Uniprot code: O31675), E. coli (Uniprot code: E2QH39), Archeoglobi sulfaticallidus (N0BDM1), Pectobactum atrosepticum (Q6D820), Geobacillus kaustrophilus (Uniprot code:GK1520), B.clausii (Q5WG41), S. rimosus (SrToyM) (Uniprot code: B6CWJ6) and S. diastatochromogenes (A0A1L7B5M4).Conserved hydrophilic residues are shown in yellow.

Twenty-five hydrophilic residues were found to be conserved across species. Some of these residues should be involved in CDG/ADG binding.

3.8.2 Finding conserved residues that are not involved in MgATP or Zn^{2+} binding

A second sequence analysis sought to identify motifs within the NS family. Such motifs should be well-conserved residues and involved in the binding of MgATP and Zn^{2+}. It is known already that MgATP binds through a phosphate-binding loop present in residues 9-14. Besides, the Zn^{2+} binding site is known to be composed of 4 cysteines (Cys186, Cys195, Cys198 and Cys201). These residues, while well conserved, should not be involved in the binding of CDG/ADG. To ensure that no other motifs are present on NSs sequences, a motif search was done on the Pfam database using MOTIF. It was found, on all the NSs tested that the residues corresponding to residues 1 to 61 on BsQueC sequence are part of a PF01171 motif, involved in the binding of ATP. For this reason, both residues 1 to 61 and the four cysteines involved in zinc binding are thought not to be involved in CDG/ADG binding. This leaves 12 residues possibly involved in CDG/ADG binding, as shown in Figure 3.64.

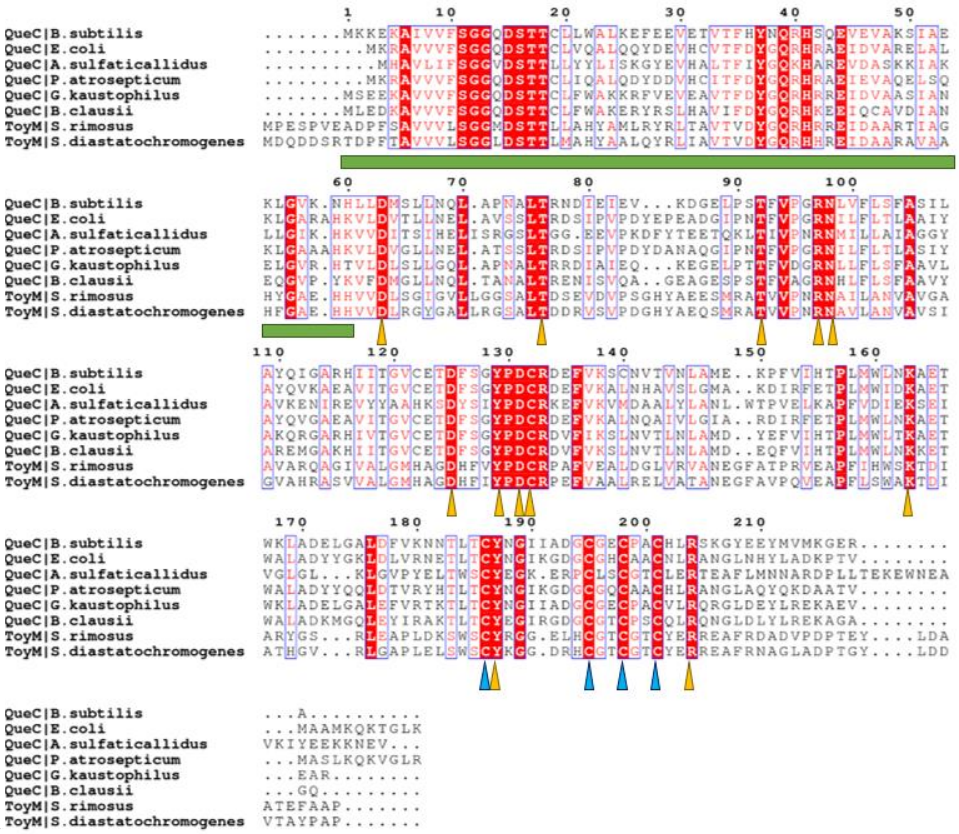


Figure 3.64: Sequence alignment the NSs from B. subtilis (BsQueC) (Uniprot code: O31675), E. coli (Uniprot code: E2QH39), Archeoglobi sulfaticallidus (N0BDM1), Pectobactum atrosepticum (Q6D820), Geobacillus kaustrophilus (Uniprot code:GK1520), B.clausii (Q5WG41), S. rimosus (SrToyM) (Uniprot code: B6CWJ6) and S. diastatochromogenes (A0A1L7B5M4). The region involved in ATP binding is shown in green, and cysteine residues involved in zinc binding are shown in blue.

3.8.3 Finding conserved residues in the vicinity of the ADP

The number of residues likely to be involved in CDG/ADG binding was reduced to 12. This number can be further reduced as it is postulated that the mechanism of the NS-catalysed reactions involves the formation of an adenylate intermediate (as shown previously on Figure 3.2). For this reason, it can be postulated that the reaction will take place close to the phosphate moiety of the MgATP molecule. The ADP-bound structure of BsQueC that was resolved previously in Part 3.5.1.3 can, therefore, be used to locate the residues possibly involved in CDG/ADG binding that are close to the phosphate moiety of the ADP. The structure of BsQueC showing 11 of the 12 potential residues potentially involved in CDG/ADG binding is shown on Figure 3.65 (one residue (D125 was part of a loop that was not rendered)).

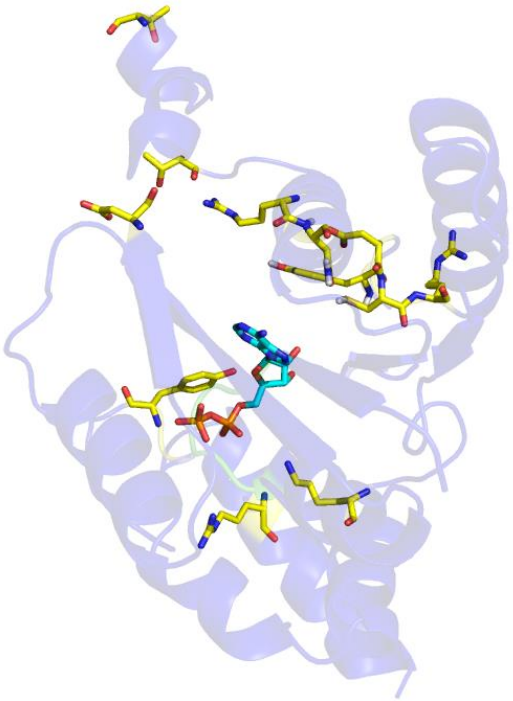


Figure 3.65: Structure of BsQueC showing 11 of the 12 residues (yellow) potentially involved in CDG/ADG binding. ADP is shown in cyan.

From Figure 3.65, three residues, Lys163, Tyr187 and Arg204, that are shown on Figure 3.66 are both close to the phosphate moiety of ADP, well conserved across all NS enzymes and also not part of any other motifs. As such, all these residues will be deactivated to investigate their role in NS catalysis.

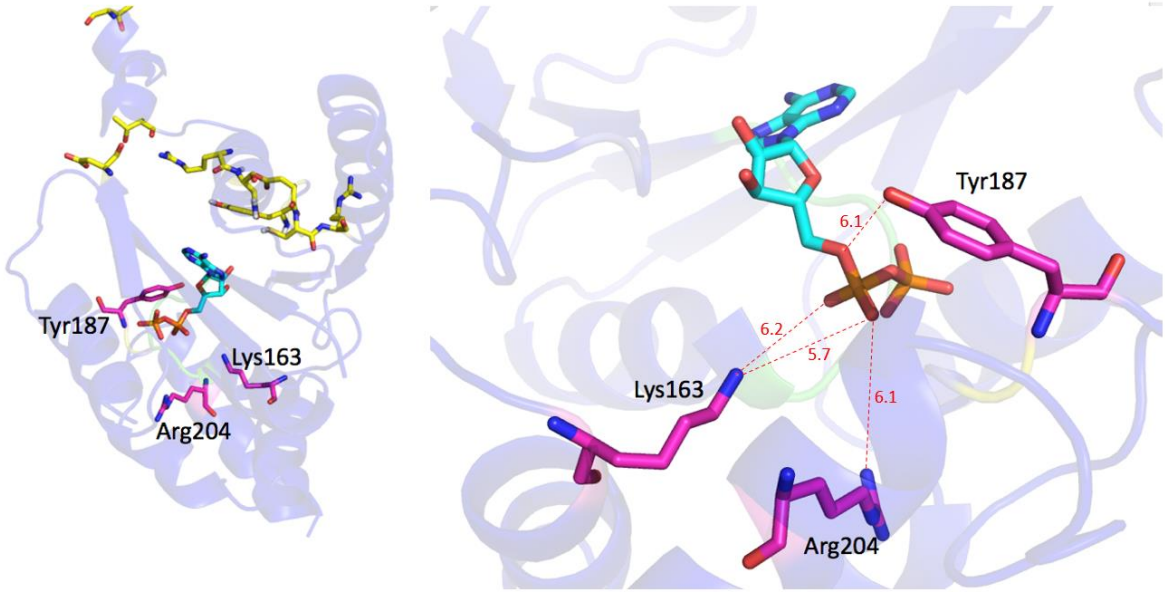


Figure 3.66: A) Structure of BsQueC highlighting the three residues (magenta) which are both conserved across all NSs, not part of any other motifs and close to ADP's phosphate moiety. ADP is shown in cyan.

As shown on Figure 3.66, intermolecular distances between the ADP molecule and the residues is slightly above the threshold for hydrogen-bonding in the current situation. However, as, in this work, ADP was co-crystallised with the enzyme, it is possible that ATP, the actual cofactor used in the NS reaction, adopts a slightly different conformation where the α-phosphate is closer to the three residues thought to be involved in the NS reaction. In addition, upon binding CDG, the conformation of the enzyme-ATP complex might also be disrupted, bringing the ATP α-phosphate within H-bond distance to these residues. Both of these hypotheses are possible considering that only a 1.5-2 Å displacement of the ATP molecule would bring it close enough to interact with the three residues. The best possible way to confirm that would be to successfully co-crystallise the enzyme with CDG and ATP. However, as discussed previously, this was attempted with no success.

3.8.4 Generation and assay of the BsQueC mutants

To assess their involvement in NS activity, these three residues were modified by site-directed mutagenesis. As such, the BsQueC K163A, Y187A and R204A one-point mutants were generated following the overlapping primers method explained in Chapter 2. These mutants were expressed and purified following the same procedure than WT BsQueC and were characterised by ESI-MS. Yields of 20-25 mg of pure enzymes per litres of culture were obtained. The ESI-MS analyses of these mutants are shown in Figure 3.67, and their theoretical (calculated with ExPaSy) and determined masses are shown in Table 3.16. The three BsQueC mutants were then assayed for the NS reaction to assess their activity, as shown in Figure 3.68 and Table 3.17.

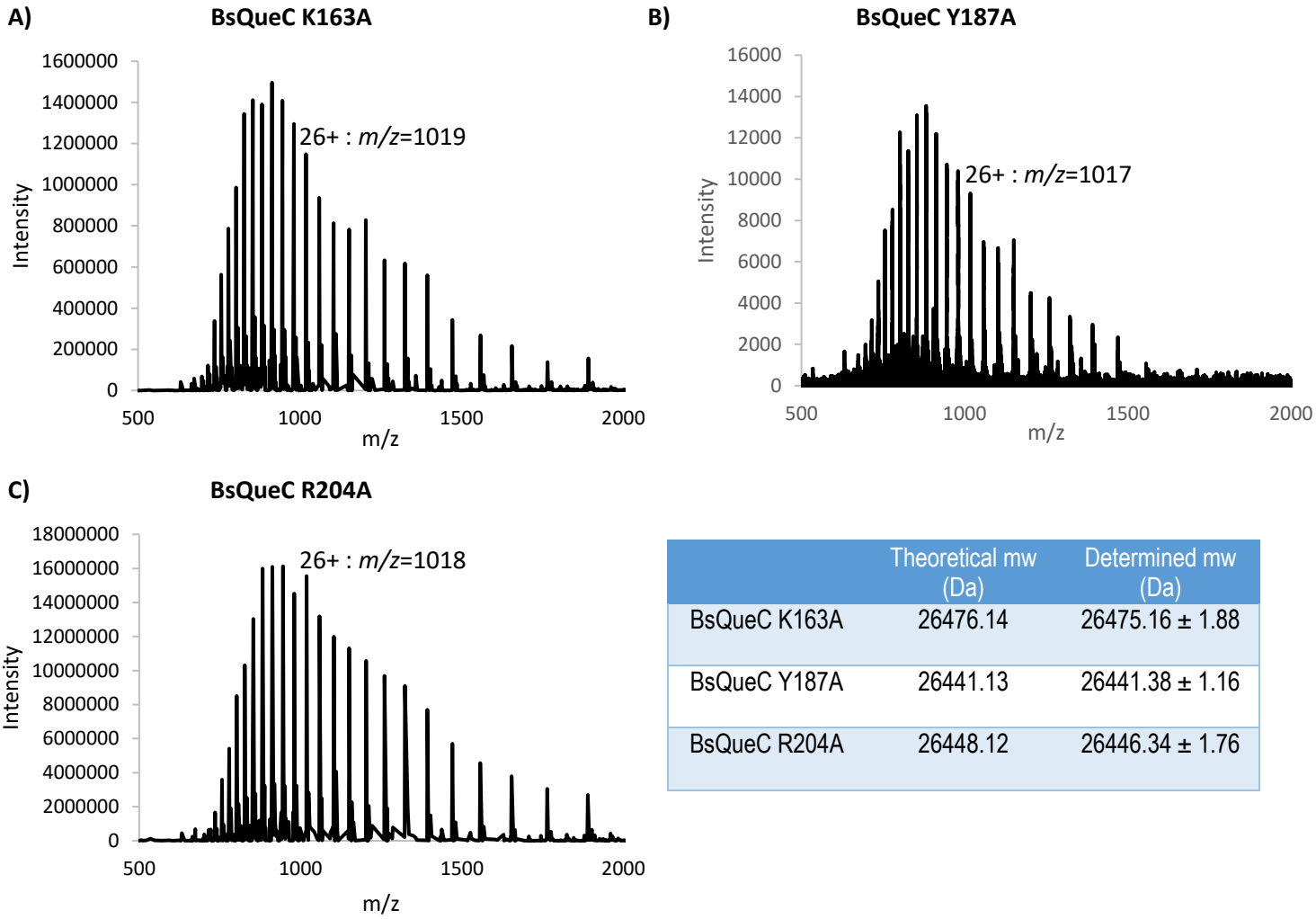


Figure 3.67: A) ESI-MS spectrum of BsQueC K163A. B) ESI-MS spectrum of BsQueC Y187A. C) ESI-MS spectrum of BsQueC R204A.

	Theoretical mw (Da)	Determined mw (Da)
BsQueC K163A	26476.14	26475.16 ± 1.88
BsQueC Y187A	26441.13	26441.38 ± 1.16
BsQueC R204A	26448.12	26446.34 ± 1.76

Table 3.16: Theoretical and experimental molecular weight for various BsQueC point mutants.

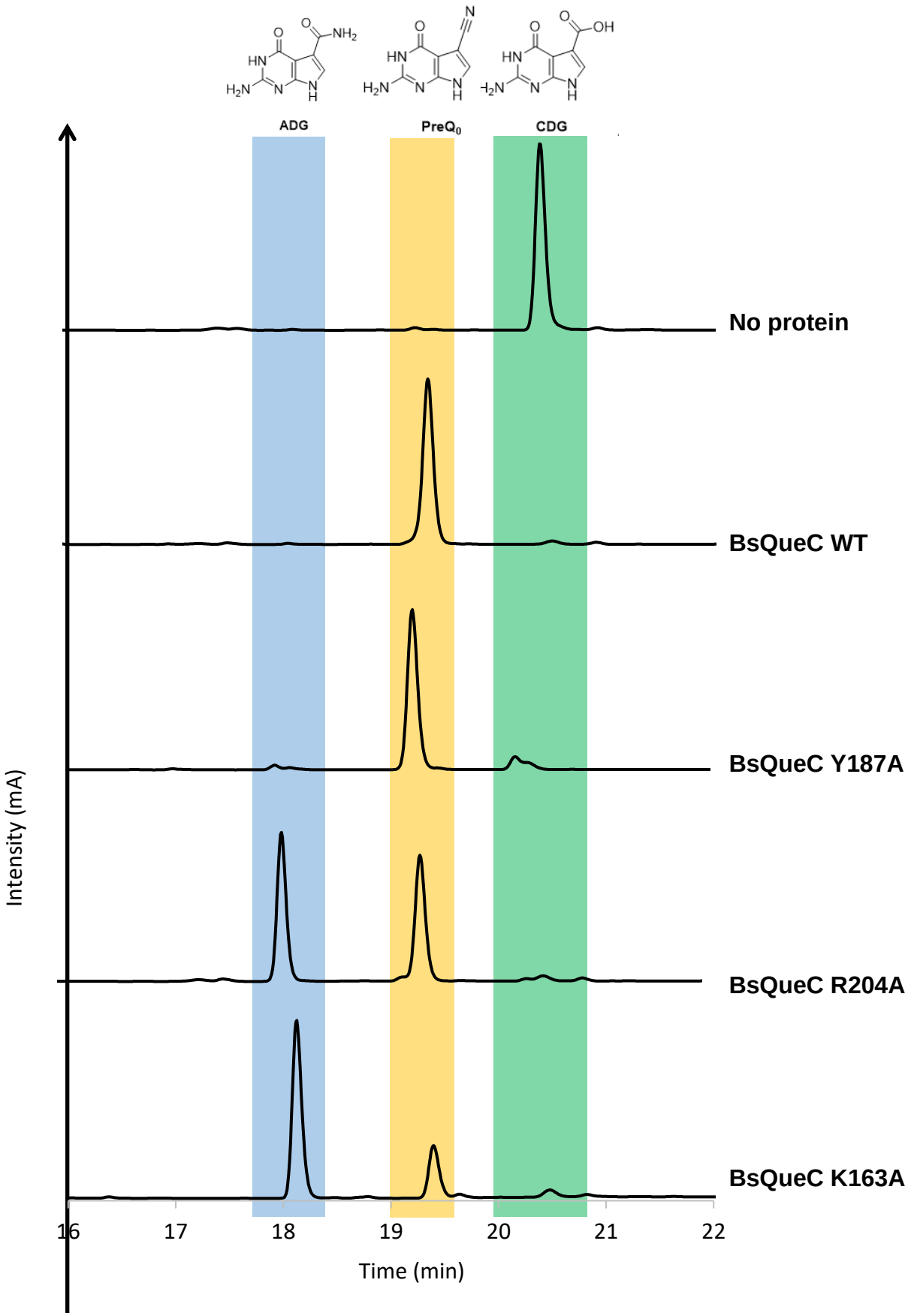
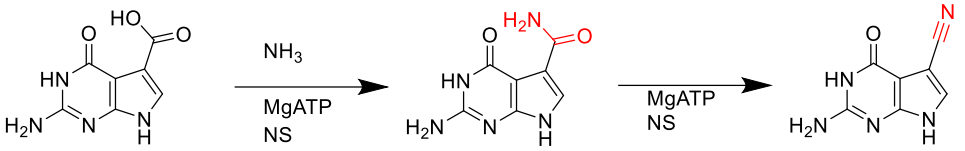


Figure 3.68: HPLC chromatograms of the NS reaction for BsQueC WT and various BsQueC point mutants after 5 hours.
Conditions: 2 mg/mL BsQueC, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.



	CDG (%)	ADG (%)	$PreQ_0$ (%)
No enzyme	100	0	0
BsQueC WT	0	0	100
BsQueC Y187A	5	3	92
BsQueC R204A	1	52	47
BsQueC K163A	1	75	24

Table 3.17: NS conversion for BsQueC WT and various BsQueC single mutants. Conditions: 2 mg/mL BsQueC, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 5 hours.

Hence, after 5 hours, 95 to 99% conversion from CDG to ADG was observed for all the BsQueC mutants. As developed earlier, for BsQueC WT, 91% conversion of ADG to $PreQ_0$ is observed. A similar conversion is observed for BsQueC Y187A, suggesting that Tyr187 is not involved in NS catalysis. However, both K163A and R204A mutations lead to significant decreases in the conversion of CDG to $PreQ_0$ in identical conditions, suggesting that both residues are involved in ADG to $PreQ_0$ conversion and are therefore involved in NS activity.

Both lysine and arginine have in common that they are positively-charged. Positively-charged residues are crucial in adenylating enzymes mechanism as they will stabilise the negatively-charged adenylate intermediate. Indeed, Schmelz and Naismith have shown that, in many but all classes adenylating enzymes, the ATP α-phosphate is surrounded by positively-charged residues (Lys and/or Arg and/or His) (Figure 3.69)[22]. This ATP α-phosphate can be considered as the focal point of the reaction, since this is where the substrate attacks upon forming the substrate-adenylate complex.

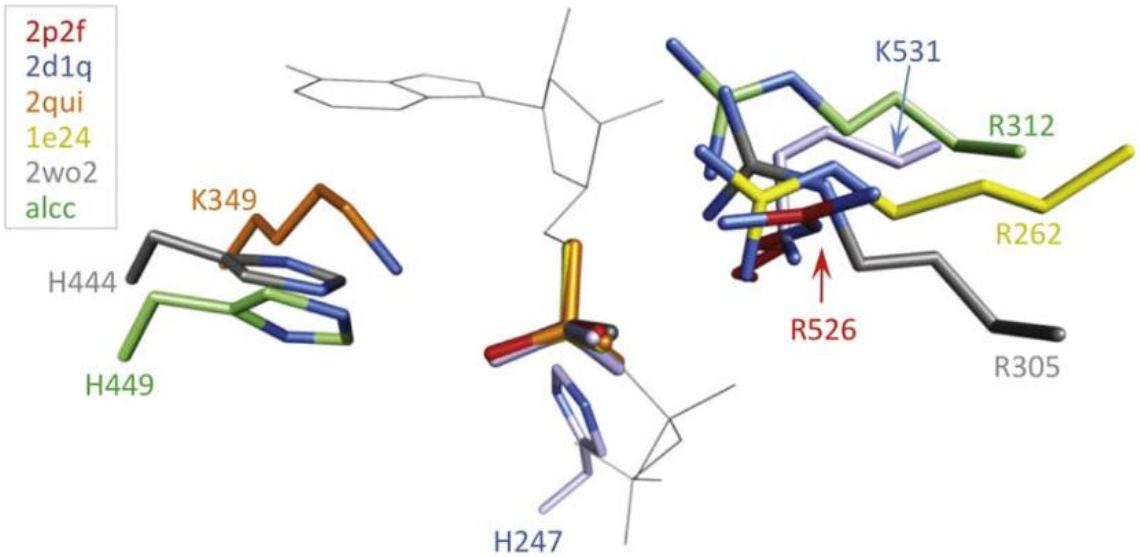


Figure 3.69: Overlay of the coordination of the ATP α-phosphate by positively-charged residues in different classes of adenylating enzymes.In all cases at least one positively-charged residue is present. Acetyl-CoA synthetase (2P2F) and Japanese firefly luciferase (2D1Q) are class I; human Trp-tRNA synthetase (2QUI) and Lys-tRNA synthetase (1E24) are class II; AcsD (2WO2) and AlcC (ALCC) are class III adenylating enzymes. Figure taken from [22].

As both Lys163 and Arg204 are positively-charged residue, this decrease of activity was expected and provides further proof of the importance of positively-charge residues in adenylate-forming enzymes. Deactivating these two positively-charged residues would disfavour the formation of the negatively-charged intermediates in the reaction. Significant effects resulting from the deactivation of a positively-charged residue in an adenylating enzyme were already observed in the research group. It was found by Tron and colleagues that the mutation of an arginine residue to glycine (R40G) in biotin protein ligase, an enzyme which catalyses the formation of biotinyl-5'-AMP from biotin and ATP, led to promisciuous biotinylation[56]. This is because the biotin adenylate is no longer "trapped" in the active site by the positively-charged arginine and can therefore escape amd react with free amines[56].

3.8.5 Generation of the double-mutant

As previous results suggest that both Lys163 and Arg204 are involved in the ADG to $PreQ_0$ conversion, the K163A R204A double mutant was generated by site-directed mutagenesis. This mutant was expressed and purified following the same procedure than WT BsQueC and was characterised by ESI-MS and SDS-PAGE. Yields of 25 mg/L of culture were obtained. The SDS-PAGE analysis and the ESI-MS spectrum of these mutants are shown in Figure 3.70.

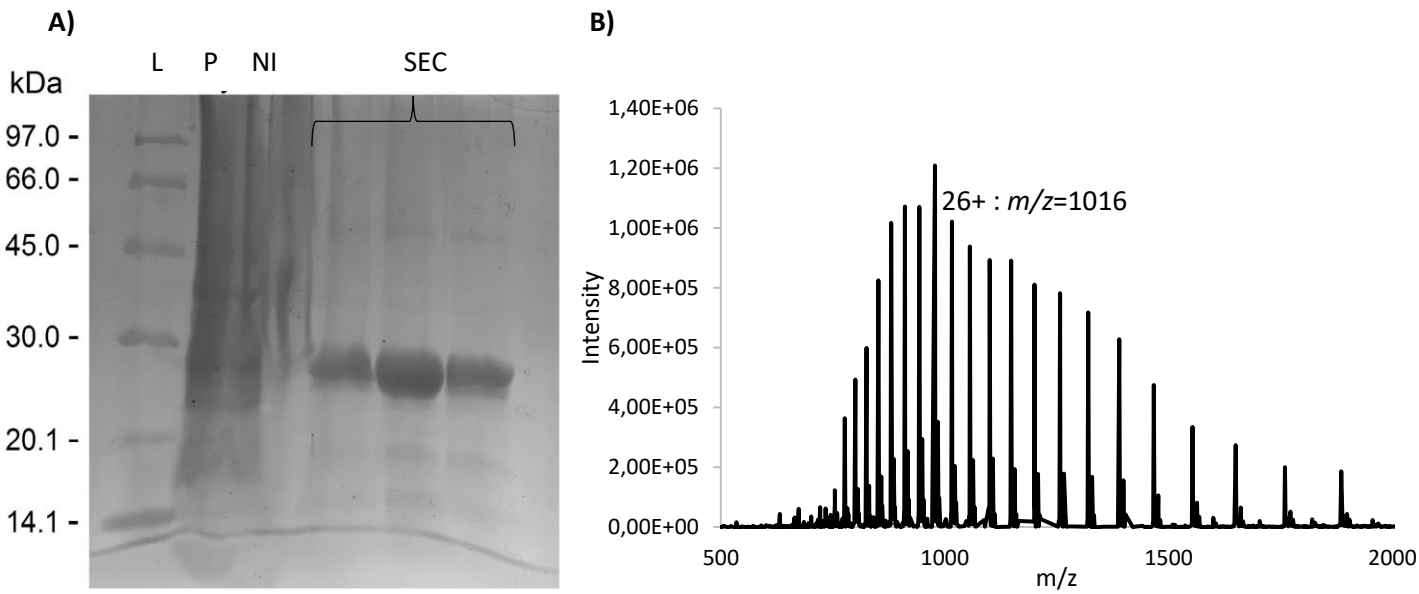


Figure 3.70: A) 12% SDS-PAGE gel for a BsQueC K163A R204A purification. L: low molecular weight ladder, P: pellets, Ni: After IMAC, SEC: after SEC.

B) BsQueC K163A R204A ESI-MS spectrum. Theoretical mass (ExPasy): 26391.03 Da; Actual mass: 26389.15 ± 2.07 Da.

.

The double mutant was assayed alongside the BsQueC WT for the NS reaction to assess its activity, as shown in Figure 3.71 and Table 3.18:

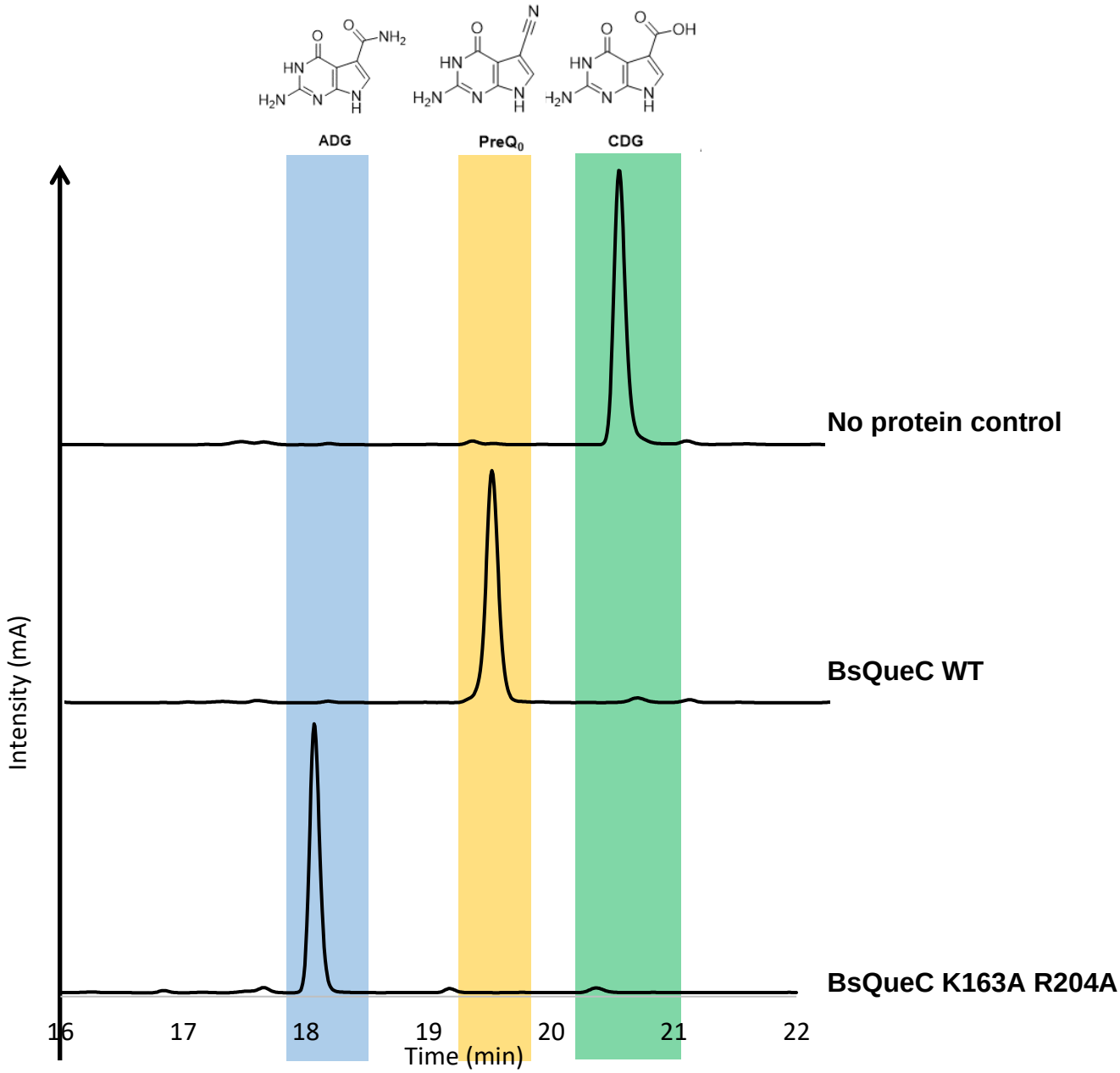


Figure 3.71: HPLC chromatograms of the NS reaction for BsQueC WT and BsQueC K163A R204A. Conditions: 2 mg/mL BsQueC, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

	CDG (%)	**ADG (%)**	**$PreQ_0$ (%)**
No enzyme	100	0	0
BsQueC WT	0	0	100
BsQueC K163A R204A	0	100	0

Table 3.18: NS conversion for BsQueC WT and BsQueC K163A R204A. Conditions: 2 mg/mL BsQueC, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 5 hours.

Hence, it was found that this double mutant NS, while retaining its amide synthetase activity, does not possess any amide dehydratase activity. It was therefore possible to separate NS two activities. This should represent the first successful engineering of an NS enzyme. This engineered amide synthetase, named BsQueC-eAmS, will be further studied and characterised below (section 3.10).

3.9 Engineering an amide synthetase from SrToyM

With the successful engineering of BsQueC accomplished, it was attempted to produce a similar AmS from SrToyM by mutating the equivalent residues on SrToyM, K176 and R213 to alanine. Hence the SrToyM K176A and R213A single-point mutants and the double-mutant were generated following the overlapping primers method explained in Chapter 2. These mutants were expressed and purified following the same procedure than for SrToyM WT. However, both SrToyM R213A and SrToyM double mutant were found to be insoluble, as shown in Figure 3.72.

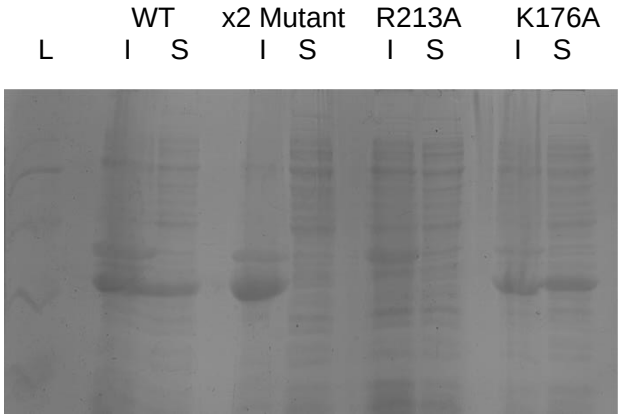


Figure 3.72: 12% SDS-PAGE gel of the cell pellets (I) and supernatant (S) of SrToyM WT, K176A-R204A (x2 Mutant), R213A and K176A. This shows that the enzyme is only soluble in the case of SrToyM WT and SrToyM K176A.
L: low molecular weight ladder

This suggests that R213 has an important role in ToyM and may be involved in a key solvent interaction that is not permitted upon its deactivation by mutagenesis. Only the SrToyM K176A single mutant could be purified with a yield of 19 mg/L of culture and characterised by ESI-MS, as shown in Figure 3.73. This mutant was also assayed, and full conversion of CDG was observed to ADG (88%) and $PreQ_0$ (12%).

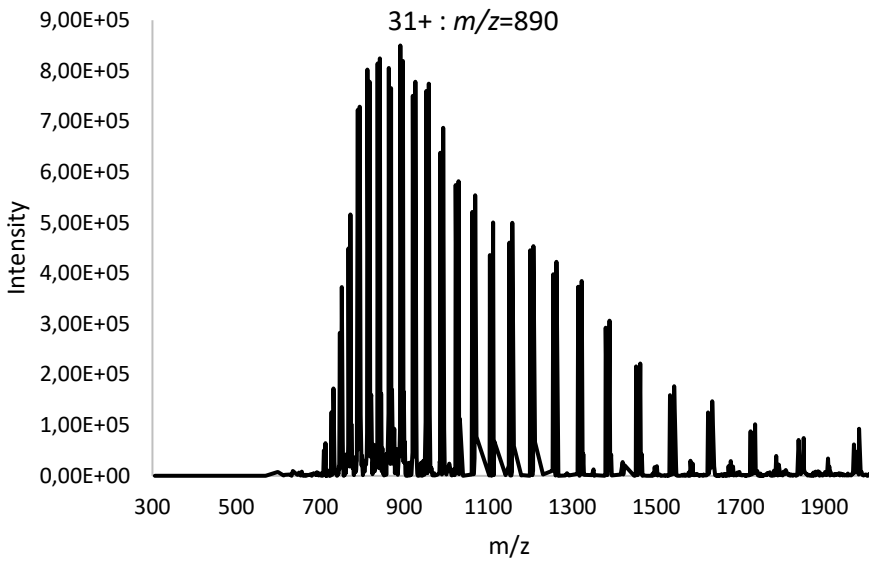


Figure 3.73: ESI-MS spectrum of ToyM K176A. Theoretical mass is 27571.80 Da. Experimental mass is 27572.92 ± 1.88 Da.

Small scale expression of the SrToyM K176A R213A with additives thought to enhance protein solubility were carried out. Leibly and coworkers published a method paper where they screened a wide range of salts, detergents, PEGs or sugars to solubilise some insoluble proteins[57]. Some of the additives which were the most successful in facilitating protein solubility were trialled. These included 700 mM trehalose, 500 mM mannitol and 400 mM L-arginine. The solubility of the protein expressed in these small scale expression tests was assessed by SDS-PAGE, as shown in Figure 3.74.

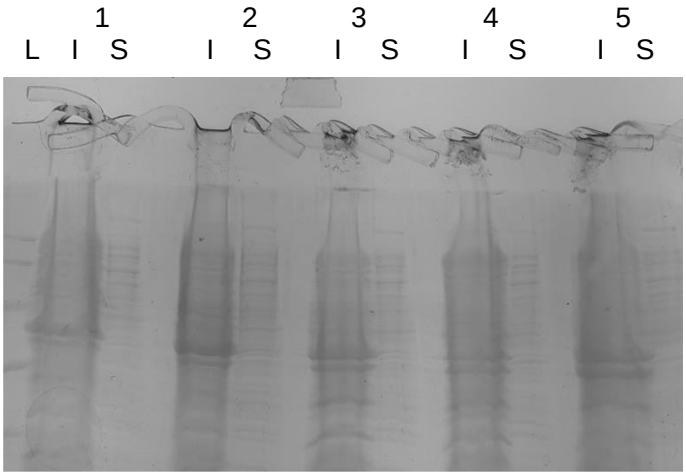


Figure 3.74: 12% SDS-PAGE gel of the cell pellets (I) and supernatant (S) of the expression of SrToyM K176A R213A with various additives. This shows that the enzyme is insoluble in all cases.
L: low molecular weight ladder
1: no additive
2: 700 mM trehalose
3: 500 mM mannitol
4: 400 mM L-arginine
5: 700 mM trehalose, 500 mM mannitol and 400 mM L-arginine

Hence, while no conditions in which SrToyM K176A R213A was soluble were found, this experiment highlighted the importance of the arginine residue in NSs behaviour. The focus was then placed on the further characterisation of BsQueC engineered amide synthetase (BsQueC-eAmS).

3.10 Characterisation of BsQueC-eAmS

3.10.1 BsQueC-eAmS kinetic studies

First, a kinetic study of the BsQueC-eAmS was carried out for the transformation of CDG to ADG by monitor production by HPLC (see Fig. 3.68). The scheme, kinetics curve and Michaelis-Menten analysis are shown in Figure 3.75, and the calculated kinetic parameters are shown in Table 3.19.

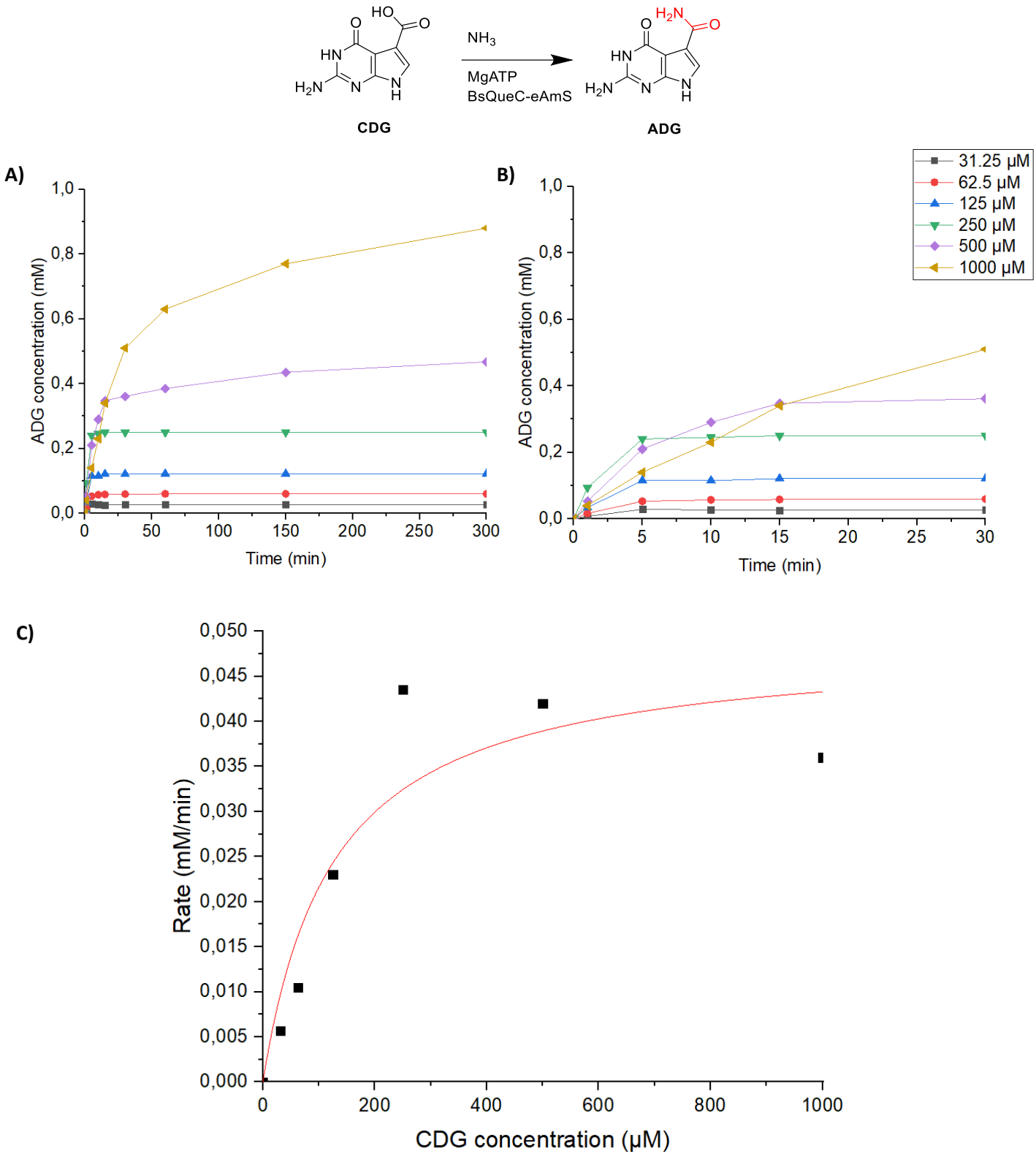


Figure 3.75: A) BsQueC-eAmS-catalysed formation of ADG monitored over time (HPLC).
B) BsQueC-eAmS-catalysed formation of ADG monitored over time (first 30 min).
C) Michaelis-Menten curve for the formation of ADG.
Conditions: 1 mg/mL BsQueC-eAmS, 0-1 mM CDG, 0-2 mM ATP, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

Parameter		Determined values
Michaelis constant	**K_M**	125 ± 6.8 µM
Turnover number	**k_{cat}**	1.27 ± 0.21 min^{-1}
Specificity constant	**k_{cat}/K_M**	1.01E-2 ± 2.70E-4 $\mu M^{-1}.s^{-1}$

Table 3.19: Kinetic parameters for BsQueC-eAmS-catalysed amide coupling between CDG and ammonium sulfate.

Hence, BsQueC-eAmS was found to have a K_M in the micromolar range and a fairly small turnover, with slightly more than one catalytic cycle per minute. This is the first time that the kinetic parameters of a NS enzyme have been determined and also, the first time a K_M for CDG was found. Interestingly, the initial rates of the 250 µM, 500 µM and 1000 µM were found to follow a negative trend, suggesting a slight product inhibition at higher concentrations. This kinetic study was also used to compare the evolution of the relative percentages of CDG, ADG and $PreQ_0$, as shown on Figure 3.76.

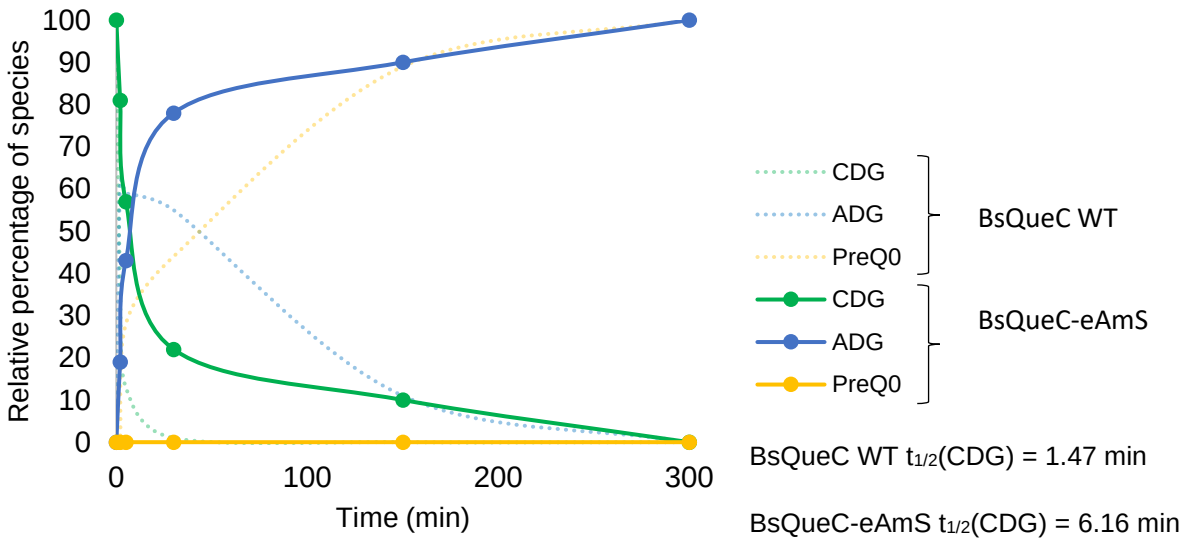


Figure 3.76: Evolution of the relative percentages of CDG, ADG and $PreQ_0$. Results for BsQueC-eAmS are shown as a bold line. Results for BsQueC WT are shown as a dashed lane. Conditions: Conditions: 2 mg/mL BsQueC-eAmS, 0.5 mM CDG, 1 mM ATP, 50 mM magnesium chloride, 50 mM amine in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

Hence, while it has already been determined that the conversion from ADG to $PreQ_0$ is completely knocked out for BsQueC-eAmS, it is interesting to note that for BsQueC-eAmS, the conversion from CDG to ADG is also slower than for BsQueC WT. While, for BsQueC WT, CDG has been fully consumed

after 30 minutes, the full conversion of CDG only occurs after 5 hours for BsQueC-eAmS. In addition, the half-life of CDG in BsQueC WT reaction is 1.47 min, while this half-life is 6.16 min when reacted with BsQueC-eAmS. This strongly suggests that the two residues that were mutated to generate BsQueC-eAmS might have a certain degree of involvement in the generation of the amide from CDG. Additionally, this also suggests that both the conversion from CDG to ADG and the hydration of ADG into $PreQ_0$ occur in a single active site. Therefore, this sheds light on the fact that the NS active site is fit to catalyse two reactions.

3.10.2 BsQueC-eAmS structural study

Structural studies were carried for the BsQueC-eAmS. The unbound structure of the BsQueC-eAmS was to be generated in order to compare the fold of BsQueC-eAmS with BsQueC WT ligand-free structure determined by Cicmil and coworkers[10]. Besides, the ligand-bound (CDG and MgATP) structure of BsQueC-eAmS was also aimed. The conditions used to obtain the ATP-bound BsQueC structure were trialled but did not yield any crystals. Consequently, HT screens were carried out for both the unbound BsQueC-eAmS (in Newcastle, in collaboration with Dr Jon Marles-Wright and Edinburgh), ATP-bound BsQueC-eAmS and CDG-bound BsQueC-eAmS (both in Edinburgh). Only one condition yielded crystals which could be further optimised, from the screens done in Newcastle with the unbound BsQueC-eAmS, as shown in Figure 3.77.

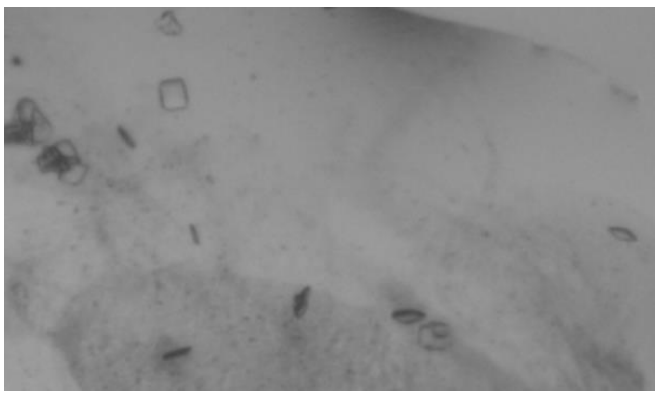


Figure 3.77: HT condition screens hit for BsQueC-eAmS.
Conditions: BsQueC-eAmS: 7.5 mg/mL; 0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate pH = 6.5, 18% w/v PEG 8000.

However, optimisation of these conditions did not yield any diffraction-quality crystals. This suggests that the mutations might either impact the stability of the protein or lie at key crystal interfaces (for example interaction with a key water molecule). In addition, this could also be due to an increased flexibility of the double-mutant due to the loss of a key interaction in the active site. In all cases, this demonstrates the importance of these two residues.

3.10.3 ADG synthesis using the BsQueC-eAmS biocatalyst

It was found, in Part 3.3.2, that ADG was hard to synthesize chemically. For this reason, this novel BsQueC-eAmS was used to synthesize this molecule at preparative scale (200 mg of starting material, split in three (66.7 mg per reaction)). To minimise the amount of enzyme required, the CDG concentration and BsQueC-eAmS loading were optimised, as shown in Table 3.20 and 3.21. HPLC chromatograms of the optimisation of the enzyme loading are shown on Figure 3.78.

CDG concentration (mM)	ATP concentration (mM)	Conversion to ADG (%)
0.5	1	100
1	2	100
3	6	74
5	10	0*

*Table 3.20: Condition screen – CDG concentration. Conditions taken forward are shown in yellow. *Precipitation of the mix was observed upon adding ATP*
Conditions: BsQueC-eAmS loading: 2 mg/mL, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 5 hours.

BsQueC-eAmS concentration (mg/mL)	Conversion to ADG (%)
2	100
1	100
0.5	98
0.25	95
0.125	70
0.0625	59

Table 3.21: Condition screen – BsQueC-eAmS loading. Conditions taken forward are shown in yellow.
Conditions: CDG: 1 mM, ATP: 2 mM, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C. All reactions were run in duplicates. The conversion was measured by HPLC after 5 hours.

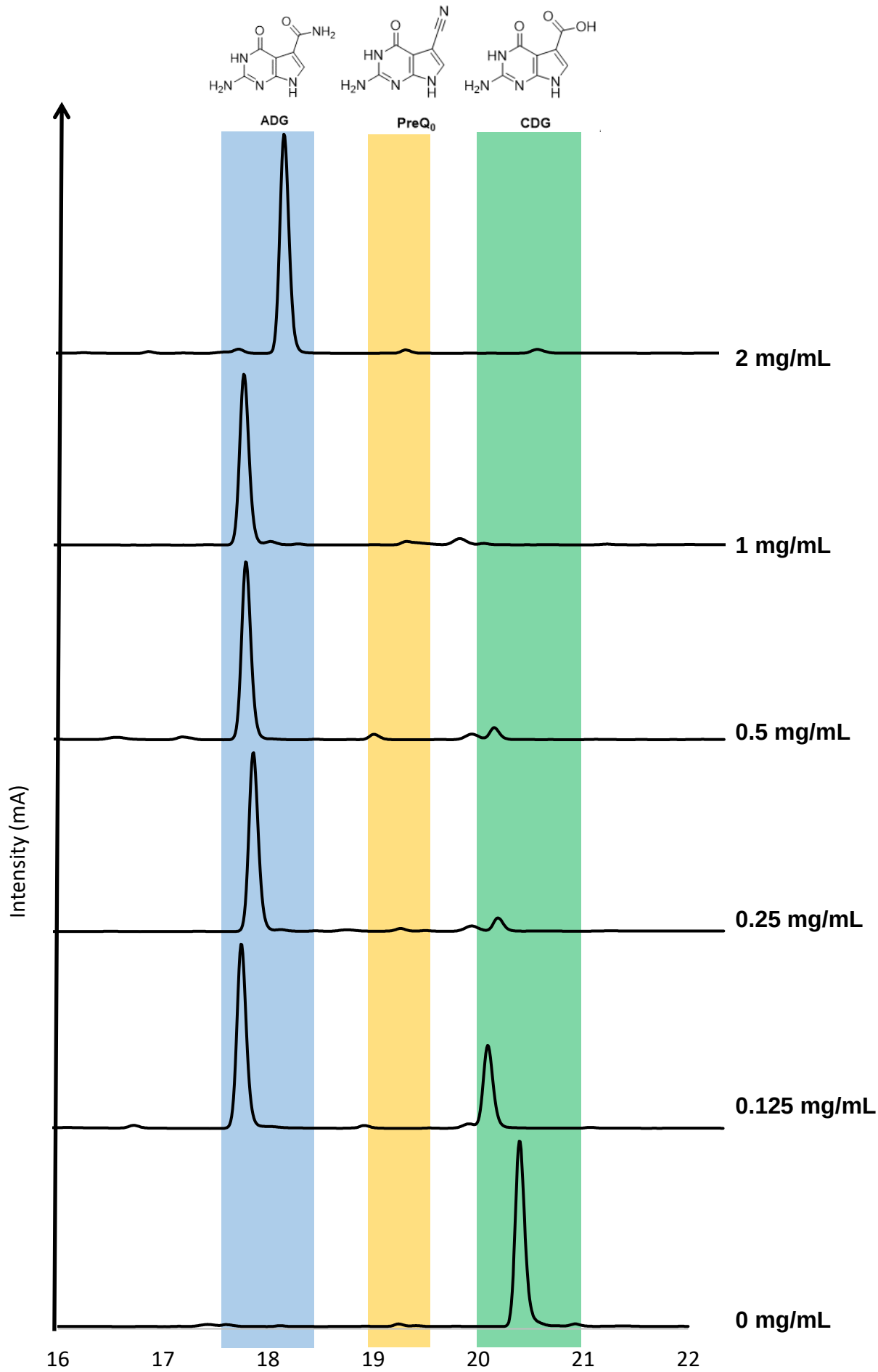


Figure 3.78: Reaction chromatograms after 5 hours for the BsQueC-eAmS reaction with multiple concentrations of enzyme.
Conditions: 2 mg/mL BsQueC, 2 mM CDG, 50 mM magnesium chloride, 50 mM ammonium sulfate in 100 mM TRIS (pH 7.0), 1 mM DTT at 37°C.

Consequently, from requiring 2 mg/mL enzyme to convert 0.5 mM CDG, only 0.25 mg/mL BsQueC-eAmS are required to convert 1 mM CDG after this optimisation. Therefore, to carry out the biotransformation planned of 200 mg CDG, the amount of enzyme required previously was 4120 mg. Based on protein yields of 25 mg of pure enzyme per litre of culture routinely obtained, 166 litres of culture would have been required to make sufficient amounts of BsQueC-eAmS to carry out this transformation. This would have been very hardly feasible in the laboratory. After optimisation, the amount of BsQueC-AmS required to carry out the same biotransformation is reduced to 258 mg. Hence, only 10.3 litres of culture was needed to produce this amount of enzyme. The reaction was carried out in three 1L conical flasks and monitored by HPLC. Interestingly, after 2 hours, full conversion of CDG was observed on all three replicates. Purification of only one of the replicates was carried (67 mg of starting material). The product was filtered and recovered alongside some precipitated BsQueC-eAmS. The product was then redissolved in 20 mL DMSO, diafiltered using a 10000 Da cut-off membrane to ensure that no precipitated enzyme was present. DMSO was then removed under high vacuum for five days. The product was finally washed with THF and recovered by filtration. 9 mg of ADG were recovered and found to be 95+% pure by HPLC and NMR. This represents the first successful direct biosynthesis of ADG from CDG, with an isolated yield of 14%. Interestingly, on the ^{1}H NMR, it was noted that the two protons of the primary amide have different chemical shifts, with a difference of about 2 ppm between the two shifts. This strongly suggests that one of the two protons of the primary amide undergoes intramolecular H-bonding interaction with the deazaguanine carboxyl, as shown in Figure 3.79. This gives extra credit on the hypothesis that the amide is able to attack the ATP α-phosphate because at least one of the amide protons acts as a hydrogen-bond donor, and therefore receive electron density, hence activating the amide for attack. The HPLC and NMRs (^{1}H and ^{13}C) of the product are shown in Figure 3.80.

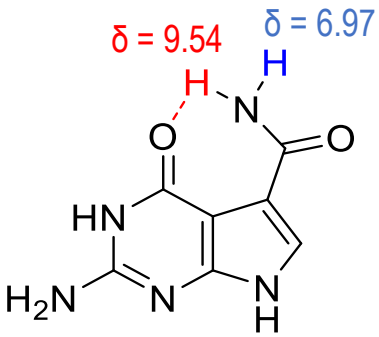


Figure 3.79: Intramolecular H-bonding interaction between one of the primary amide hydrogen atoms and the deazaguanine carboxyl, explaining the ^{1}H NMR shift difference between these two protons

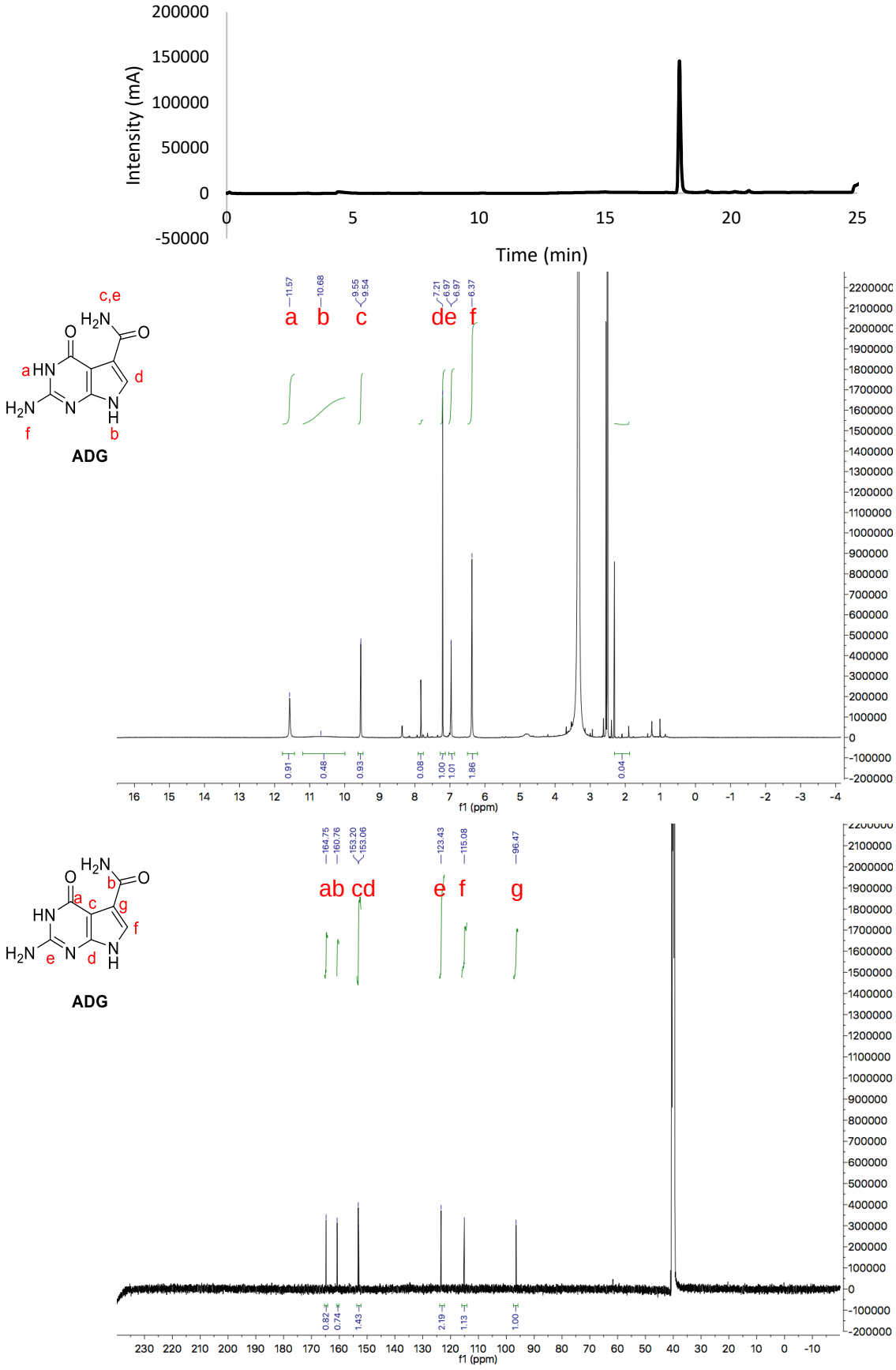


Figure 3.80: A) HPLC chromatogram of the ADG produced by BsQueC-eAmS. One single peak (retention time: 18 min), corresponding to ADG, is observed.
B) ^{1}H NMR spectrum (600 MHz, DMSO)
C) ^{13}C NMR spectrum (600 MHz, DMSO)

3.11 Conclusion

This work explored the structure, substrate specificity and mechanism of the NS biocatalysts. Two recombinant NSs, MgATP-dependent BsQueC and SrToyM were successfully expressed and purified in good yields and purities. A robust HPLC assay was then developed and used to assay these enzymes for their natural NS reaction: the transformation of CDG into the corresponding nitrile, $PreQ_0$, via the primary amide ADG, as shown on Figure 3.81.

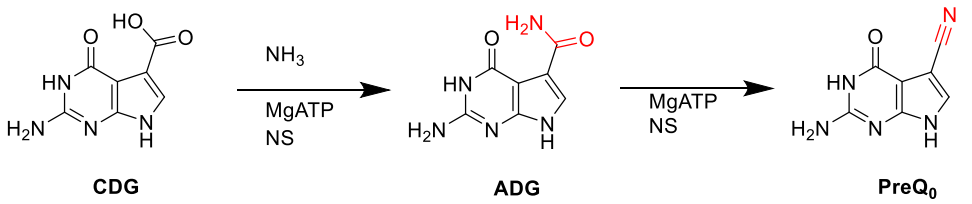
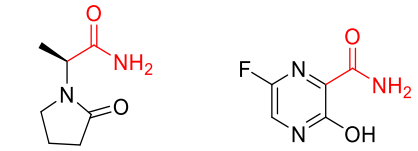


Figure 3.81: NS reaction scheme. CDG is first transformed into ADG, using one equivalent of MgATP and then into $PreQ_0$ using a second equivalent of MgATP.

It was found that, while complete conversion of CDG into $PreQ_0$ was observed for BsQueC, a mixture of ADG and $PreQ_0$ was obtained with SrToyM. The substrate scope of both biocatalysts was then explored, and it was confirmed that both NSs have a very narrow acid substrate scope, only accepting their natural ligand and close derivatives. However, it was found that a fairly wide range of CDG-derived secondary and tertiary amides could be produced using both NS enzymes tested. A rigorous structural study of these NS enzymes was then carried out, leading to the resolution of the ADP-bound BsQueC structure with a resolution of 2.1 Å. This structure was the cornerstone of the rational engineering carried on BsQueC to engineer an amide synthetase from the NS (BsQueC-eAmS) and to separate the two activities. Three residues (K163, Y187 and R204) were identified and mutated to alanine, and it was observed that two of these mutations (K163A, R204A) were important as significant losses in amide dehydratase activity were observed. This suggests that these residues are not involved in the first step of the reaction, rather they play a role in the second amide to nitrile conversion. Generating a double mutant, which combined both mutations, finally led to the complete disappearance of the amide dehydratase activity, allowing this engineered enzyme to join the amide synthetase family. However, as an impact on the speed of the conversion from CDG to ADG was observed, these observations weaken the hypothesis that the two reaction steps proceed in two different active sites. This novel amide synthetase was further characterised through kinetic and structural studies. For the first time, the kinetic parameters for CDG were determined. In addition, this enzyme was used to carry out the first direct biosynthesis of ADG.

Further development of this project would be to engineer this novel amide synthetase to expand its substrate scope and allow it to produce industrially-relevant primary amides. Examples could include Levetiracetam, an anticonvulsant drug used to treat epilepsy and the 60th top-selling small molecule in 2018[58], or Favipiravir, an antiviral drug which was the first molecule approved in China to treat the 2019 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[59]. Their structures are shown in Figure 3.82.



Levetiracetam **Favipiravir**

Figure 3.82: Structures of Levetiracetam and Favipiravir

In addition, as the residues involved in the amide dehydratase activity of the enzyme have been uncovered, saturation mutagenesis of these residues could be carried to engineered, enhanced nitrile synthetase enzymes.

References

[1] M. Winkler, K. Dokulil *et al.,* 2015, *Chem.Biol.Chem.,* 16, pp 2373-2378

[2] Y.C. Chen, V.P. Kelly *et al.,* 2010, *RNA,* 16, pp 958-968

[3] M. Vinayak and C. Pathak, 2009, *Biosci.Rep,* 30, pp 135-148

[4] T. Uematsu and R.J. Suhadolnik, 1970, *Biochen.,* 9 pp 1260-1266

[5] Y. Becker, 1980, *Pharmac.Ther.,* 10, pp 119-159

[6] X.-P. Shentu, Z.-Y. Cao *et al.,* 2018, *PLOSOne,* 13, e0203006

[7] M. Ri, E. Tashiro *et al.,* 2012, *Blood Cancer J.,* 2, e79

[8] T. Ploom, C. Haussmann *et al.,* 1999, *Structure,* 7, pp 509-516

[9] G. Schoedon, U. Redweik and H.C. Curtius, 1989, *Eur.J.Biochem.,* 178, pp 627-634

[10] R.M. McCarty, A. Somogyi and V. Bandarian, 2009, *Biochem.,* 48, pp 2301-2303

[11] C.R. Loomis and R.M. Bell, 1988, *J.Biol.Chem.,* 263, pp 1682-1692

[12] M.T. Nelp, A.V. Astashkin *et al.,* 2014, *Biochem.,* 53, pp 3990-3994

[13] M.T. Nelp and V. Bandarian, 2015, *Angew.Chem.Int.Ed.,* 54, pp 10627-10629

[14] R.M. McCarty, A. Somogyi *et al.,* 2009, *Biochemistry,* 48, pp 3847-3852

[15] N. Cicmil and R.H. Huang, 2008, *Proteins,* 72, pp 1084-1088

[16] D.G. Brown and J. Boström, 2014, *J.Med.Chem.,* 59 pp 4443-4448

[17] P.N. Devine, R.M. Howard *et al.,* 2018, *Nature Rev.Chem.,* 2, pp 409-421

[18] D. Liu, Y.L. Guo *et al.*, 2018, *Beilstein J.Org.Chem.,* 14, pp 1112-1119

[19] A. Isodoro-Llobret, M.N. Kenworthy *et al.,* 2019, *J.Org.Chem*, 84, pp 4615-4628

[20] M. Funabashi, Z. Yang *et al.,* 2010, *Nat.Chem.Biol.,* 6, pp 581-586

[21] M. Winn, S.M. Richardson *et al.,* 2020, *Curr.Opin.Struct.Biol.,* 55, pp 77-85

[22] S. Schmelz and J.H. Naismith, 2009, *Curr.Opin.Struct.Biol.,* 19, pp 666-671

[23] M.T. Miller, B. Gerratana *et al.,* 2003, *J.Biol.Chem.,* 278, pp 40996-41002

[24] A.M. Gulick, 2009, *ACS Chem.Biol.,* 4, pp 811-827

[25] S. Schmelz, N. Kadi *et al.*, 2009, *Nat.Chem.Biol.*, 5, pp 174-182

[26] M. Wang, L. Moynié *et al.,* 2017, *Nat.Chem.Biol.*, 13, pp 660-667

[27] M. Petchey, A. Cuetos *et al., Angew.Chem.Intl.Ed.*, 57, pp 11584-11588

[28] P.M. Marchetti, S.M. Richardson *et al.,* 2019, *Med.Chem.Commun.,* 10, pp 1192-1196

[29] T. Tsuda, M. Asami *et al.*, 2014, *Biochem.*, 53, pp 2650-2660

[30] E.M. Gabor and D.B. Janssen, 2004, *Prot.Eng.Design.Select.,* 17, pp 571-579

[31] A.B. Cuthbertson, A.D. Rodman, 2019, *Processes,* 7, pr7060318

[32] Global Newswire, accessed from https://www.globenewswire.com/news-release/2019/11/04/1940514/0/en/Amoxicillin-Market-To-Reach-USD-4-962-6-Million-By-2026-Reports-And-Data.html on 11th December 2019

[33] GlaxoSmithKline Intellectual Property Development Limited, 2016, Patent WO16193255 A1

[34] M. Sato, Y. Masuda *et al.,* 2007, *J.Biosci.Bioengineering,* 103, pp 179-184

[35] S.M. Resnick and A.J.B. Zehnder, 2000, *App.Environm.Microbiol.,* 66, pp 2045-2051

[36] J.I. Ramsden, R.S. Heath *et al.,* 2019, *J.Am.Chem.Soc.,* 141, pp 1201-1206

[37] H. Cao, C. Li *et al.,* 2018, *App.Biochem.Biotechnol.,* 185, pp 385-395

[38] E.A. Letts, 1872, *Proceed.Roy.Soc.London,* 21, pp 139-147

[39] A. Kekule, 1872, *Berichte der Deutschen Chemischen Gesellschaft,* 5, pp 669-674

[40] R. Mitchell, 1931, *J.Am.Chem.Soc.,* 53, pp 327-331

[41] A. Oroskar, P. Kurek and A. Oroskar, 2011, US Patent US2011124901 A1

[42] S. Galanov, 2014, *Pet.Chem.*, 2014, 54, pp 387-391

[43] F. Juncai, B. Liu *et al.*, 1996, *Synth.Commun.*, 26, pp 4545-4548

[44] N.K. Bhattacharyya, S. Jha *et al.,* 2012, *Intl.J.Chem.App.,* 4, pp 295-304

[45] S.A. Shipilovskikh, V.Y. Vaganov *et al.,* 2018, *Org.Lett.,* 20, pp 728-731

[46] M.T. Migawa, J.M Hinkley *et al.*, 2006, *Synth.Commun.,* 26, pp 3317-3322

[47] G. Zhu, Z. Liu *et al.,* 2008, *Heterocycles,* 75, pp 1631-1638

[48] J. Jung and B. Nidetzky, 2018, *J.Biol.Chem.,* 293, pp 3720-3733

[49] A. Mohammad, A. Bon Ramos *et al.,* 2017, *Biomol.,* 7, pp 30-43

[50] B.W. Lee, S.G. Lanen *et al.,* 2007, *Biochem.,* 46, pp 12844-12854

[51] M.R. Webb, 1992, *Proc.Natl.Sci.USA,* 89, pp 4884-4887

[52] T.P. Geladopoulos, T.G. Sotiroudis and A.E. Evangelopoulos, 1991, *Anal.Biochem.,* 192, pp 112-116

[53] J. Jancarik and S.H. Kim, 1991, *J.Appl.Cryst.,* 24, pp 409-411

[54] A.J. McCoy, R.W. Grosse-Kunstleve *et al.,* 2007, *J.Appl.Cryst.,* 40, pp 658-674

[55] M. Kallberg, H. Wang, *et al.,* 2012, *Nature Protocols,* 7, pp 1511-1522

[56] C.M. Tron, I.W. McNae *et al.,* 2009, *J.Mol.Biol.,* 387, pp 129-146

[57] D.J. Leibly, T.N. Nguyen *et al.,* 2012, *PLoS ONE,* e52482

[58] N.A. McGrath, M. Brichacek and J.T. Njardarson, 2010, *J.Chem.Ed.,* 87, pp 1348-1349

[59] Chinese Clinical Trials Registry (ChCTR), ChiCTR2000029600, accessed on 28th March 2020

Chapter 4 Summary and outlook

Before designing enhanced biocatalysts and carrying out directed evolution, it is important to get an understanding of the mechanism of action of the enzyme. Where and how to "hit"? The aim of this thesis was to understand the catalytic machinery involved in two biocatalysed reactions: the EcBioH-catalysed MBH reaction and the NS-catalysed ATP-dependent conversion of carboxylic acids into the corresponding nitrile.

As such, firstly, EcBioH was successfully cloned, expressed, purified and characterised. A robust HPLC assay was developed along with efficient reaction conditions for the MBH coupling of 4-NBA and MVK. It was also found that no enantioselectivity of the MBH product is obtained upon using EcBioH. A search of the residues involved in MBH catalytic activity was then performed. The catalytic triad (S82, D207, H235) involved in EcBioH esterase activity was first targeted. At first, these residues, along with the serine of another, cryptic, catalytic triad (S55) were deactivated by site-directed mutagenesis. No decrease in MBH activity was observed, suggesting that these triads are not involved in MBH catalysis. The involvement of a catalytic cysteine was also tested but it was also found that a catalytic cysteine is unlikely to be involved in MBH catalysis. The activity of EcBioH was then compared to one of another pimeloyl-ACP methyl ester esterase, HiBioG but no significant difference in MBH activity was found. Finally, the involvement of histidine was explored by expressing variants of EcBioH with different numbers of hexahistidine tags. It was found that the numbers of hexahistidine tags have a significant impact on MBH catalysis. With this result in hand, it was found that other hexahistidine tagged proteins, including denatured EcBioH and *E. coli* ACP also exhibit comparable MBH activities. It was therefore concluded that any protein that contains a hexahistidine tag could catalyse the MBH reaction. A mechanistic rationale, involving a synergic relationship between the histidine residues of the tag and the oxyanionic reactive intermediates, which could explain the increase in activity compared to imidazole, was postulated.

With these data in hand, further work could be to construct a histidine-rich active site. While a hexahistidine tag is hard to engineer due to its flexibility, having multiple histidine residues in a more rigid active site shall permit easier engineering to build enantioselectivity and increase the activity of the biocatalyst.

In a second project, NSs, a small enzyme family which converts a carboxylic acid first into the corresponding primary amide and then into the corresponding nitrile, were studied. The aim of this project was to separate these two reactions and to engineer a primary amide synthetase from NS enzymes. Two NSs, BsQueC and SrToyM, were studied. Their substrate scope was assessed and it was

found that, while they both have a very poor acid substrate scope, they are both accepting a small range of amines, which make both enzymes fit to produce secondary and tertiary amides. It was also aimed to obtain a ligand-bound structure of a NS enzyme which will be used for rational evolution of the enzyme. The structure of ADP-bound BsQueC was successfully determined and used to locate residues which may be involved in catalysis. Three residues (K163, Y187 and R204) were identified and deactivated by site-directed mutagenesis. The mutations of the two positively-charged residues (K163A and R204A) resulted in substantial reduction of the conversion from primary amide to nitrile. The double mutant, which contained both mutations was completely knockout for amide to nitrile conversion, making this mutant, *de facto,* an engineered amide synthetase. This novel primary amide synthetase was further characterised and used to produce 7-amidodeazaguanine (ADG) from 7-carboxydeazaguanine (CDG) at laboratory scale.

As strong evidence of the residues involved in NS amide hydration was obtained, further engineering could be carried with the aim of enhancing the substrate scope of the enzyme. Hence, site saturation mutagenesis could be carried on the two residues found to be involved in the reaction (Lys163 and Arg204). In this technique, the WT gene will amplified by site-directed mutagenesis PCR with a mixture of primers containing all the possible mutations into the other 19 amino acids. The mutants generated would then be assayed on other, industrially-relevant, substrates to assay their activities against non-natural ligands and all the hits, for which the mutants display some activity, will be sequenced. Further rounds of evolution, to mutate residues in the vicinity of Lys163 and Arg204 are also possible. However, it is clear that a CDG + ATP bound structure of BsQueC, SrToyM or any other NS (such as *E.coli* QueC) would be very useful to get a clearer understanding of the binding mode of the ligands. If that is not feasible, the structure generated in this work could be used to carry molecular dynamics simulations of the conversion of CDG into ADG and $PreQ_0$. It would be especially interesting to get a better understanding on how the substrate is “trapped” in the active site for both reactions.

Through generating the first ligand-bound structure of BsQueC, targeted enzyme engineering could be carried, leading to the generation a double-mutant where the complete disappearance of the amide dehydration activity was observed. Through carrying a comparative assay of the NS reaction for the WT enzyme and the mutants, it was observed that the amide synthesis step of the NS reaction was also slowed by the mutations, strongly suggesting that one single active site is responsible for both reactions catalysed by NS enzymes. As a result, as well as being the first time that both reactions catalysed by NS enzymes were isolated by enzyme engineering, this work also provides strong evidence on the presence of one single active site in NS reactions. This new data is important as that

suggests that engineering this active site could impact both reactions and, therefore, could be done to expand the substrate scope of NS enzymes for them to be fit to catalyse the syntheses of industrially-relevant amides and nitriles.

Chapter 5 Materials and Methods

5.1 Materials

All reagents, chemicals and media were purchased from Sigma-Aldrich, Thermo Fisher Scientific, Alfa Aesar, Fluorochem or Bio-Rad unless otherwise stated. All primers were purchased from Sigma-Aldrich and competent cells, pET plasmids and restriction enzymes from New England Biolabs (NEB), Novagen, Promega and Life Technologies. Protein purification columns were purchased from GE Healthcare. HPLC columns were bought from Phenomex. The codon-optimised *S. rimosus toyM* gene was purchased from GenScript.

5.1.1 Growth media

All growth media were prepared by dissolving the components listed in Table 5.1 in deionised water. The mixture was then sterilised by autoclaving for 20 min at 120°C.

Medium	Components
Lysogeny broth (LB)	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract
LB Agar	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, 15 g/L agar
Super optimal broth (SOC)	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, 2.5 mM KCl, 10 mM $MgCl_2$, 2% w/v glucose
SelenoMet™ base Medium	21.6 g/L Molecular Dimensions SelenoMet™ base medium powder
SelenoMet™ Nutrient mix*	102 g/L Molecular Dimensions SelenoMet™ nutrient mix powder

Table 5.1: Components of the growth media used in this work
**Sterilised by filtration (33 μm) instead of autoclaving*

Appropriate antibiotic was added to the cool growth media after sterilisation. Kanamycin and ampicillin stocks were made with concentrations of 30 mg/mL and 100 mg/mL respectively. The stocks were dissolved in water and were added to media in a ratio of 1:1000 for final concentrations of 30 μg/mL or 100 μg/mL.

5.1.2 Competent cell lines

Various cell lines from *E. coli* were used to store and amplify plasmids for cloning and recombinant protein expression. These are presented in Table 5.2

Cell line	Supplier	Application
BL21 (DE3)	New England Biolabs	Protein expression
C2987	New England Biolabs	High transformation efficiency and amplification during cloning
DH5α	Life Technologies	Plasmid storage and amplification
B834	Novagen	Expression of selenomethionine-enriched proteins

Table 5.2: E. coli cell lines used in this work

5.1.3 Plasmids

Target genes were cloned into various plasmids to yield recombinant proteins with specific hexahistidine tags or for cloning as shown in Table 5.3

Plasmid	Resistance	Applications
pET22b	Ampicillin	Protein expression of proteins with a C-terminal hexahistidine tag or no tag
pET28a	Kanamycin	Protein expression of proteins with a N-terminal hexahistidine tag or N- and C-terminal hexahistidine tags
pGEM T-Easy	Ampicillin	Cloning/blue-white screen

Table 5.3: Plasmids used in this work

5.1.4 Buffers

Buffers were prepared by dissolving components in deionised water and adjusting to desired pH using hydrochloric acid (HCl, 33% w/v) and sodium hydroxide (NaOH, 5 M). These were filtered through a 0.22 µm filter before chromatography. The buffers used are shown in Table 5.4.

Buffer	Components
A	50 mM Tris, pH = 7.5
B	50 mM Tris 1 M NaCl, pH = 7.5
C	20 mM tris(hydroxymethyl)aminomethane (Tris), 500 mM NaCl, 10% Glycerol 30 mM imidazole, pH = 8.0
D	20 mM tris(hydroxymethyl)aminomethane (Tris), 500 mM NaCl, 10% Glycerol 250 mM imidazole, pH = 8.0
E	20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 500 mM NaCl, 10% Glycerol pH = 7.5
F	500 mM KCl, 40 mM imidazole, pH = 7.5
G	500 mM KCl, 500 mM imidazole, pH = 7.5
H	100 mM Tris, 1 mM 1,4-Dithiothreitol (DTT), pH = 7.0

Table 5.4: Buffers used in this work

5.2 Molecular cloning

5.2.1 Overview

EcBioH-N was cloned at the beginning of this study following the procedure highlighted in Section 5.2.3. EcBioH-C and EcBioH-x have been previously cloned by Dr Menglu Wang. The EcBioH-NC clone was sub-cloned by cutting the C-terminal hexahistidine tagged EcBioH gene from a pET22a plasmid (Nde1/Xho1) and transferring it to a Nde1/Xho1 digested pET28a plasmid. BsQueC and BsQueF were previously cloned by Mari Nelson. The full length, codon optimised *toyM* (*S. rimosus*, Uniprot code: B6CWJ6) was purchased from GenScript. After cutting with Nde1/Xho1, the *toyM* gene was sub-cloned into Nde1/Xho1 digested pET28a plasmid to give a recombinant SrToyM plasmid which could be used to express SrToyM with a non-cleavable N-terminal hexahistidine tag.

5.2.2 General procedures

5.2.2.1 Gel electrophoresis and extraction

DNA gels were prepared by putting agarose (1 g) in TAE buffer containing 40 mM Tris, 20 mM acetic acid and 1 mM EDTA (100 mL). The mixture was microwaved until complete dissolution of the agarose and cooled to 60°C before adding GelRed (10 µL). The solution was stirred, poured into a casting mold and set at 4°C. PCR products and other DNA samples were mixed with 6x Purple DNA loading dye (NEB) and loaded onto the gel. A DNA ladder (Hyperladder I, Bioline) was also loaded. Gels were run at a constant voltage of 100 V for around 30 min. DNA bands were visualised using UV light and DNA bands of interest were excised from the gel using a scalpel, as required. The gel band was

weighed and the DNA was extracted using a ThermoScientific gel extraction kit according to the manufacturer's instructions.

5.2.2.2 Restriction digest

The PCR product or plasmid to be digested (30 µL) was mixed with 5* CutSmart buffer (6 µL), and the restriction enzymes (Nde1 and Xho1 2x 1.5 µL). Deionised water was added to a final volume of 50 µL . The reaction was incubated at 37°C overnight. The product was then purified by gel electrophoresis and extracted as per procedure 5.2.1.1.

5.2.2.3 Analytical digest

The plasmid to be digested (6 µL) was mixed with CutSmart buffer (2 µL), and the restriction enzymes (Nde1 and Xho1 2x 1 µL). The reaction was incubated at 37°C overnight. The product was then analysed by gel electrophoresis. Positive clones have two bands, one with correct size for backbone, one for correct size for gene of interest. Positive clones were sequenced by Eurofins Genomics using their T7 and pET-RP primers to verify that the gene of interest was correctly incorporated.

5.2.2.4 Ligation

The plasmid backbone (70 ng) and insert containing the gene of interest (~250 ng, 1:10 molar ratio of backbone to insert) were mixed with T4 DNA ligase (1.5 µL) and 10x ligation buffer (1 µL) and made up to 10 µL with deionised water. The reaction was incubated at 37°C for 3 hours. 5 µL of the reaction were then used to transform C2987 cells which were grown on LB + appropriate antibiotic.

5.2.2.5 Miniprep

Single colonies were picked from agar plates and inoculated in 10 mL LB supplemented by the appropriate antibiotic. These cultures were shaken at 37°C for 16 hours. These cultures were then centrifuged at 3500 rpm in a Hearaeus Function Line centrifuge for 15 minutes. The supernatant was decontaminated with Virkon and discarded. The plasmid was extracted from the pellet using a Qiagen QIAprep miniprep kit according to the manufacturer instructions. Plasmid DNA was eluted in deionised water instead of elution buffer provided by the kit and stored at -20°C.

5.2.2.6 Site-directed mutagenesis

The sequence of the primers used for this work are shown in Table 5.5. The overlapping method of site-directed mutagenesis developed by Liu and Naismith was used to construct the primers[1].

Protein	Primer	Sequence (5'-3')	Melting temperature (T_m) (°C)
EcBioH	**S53A For**	GGGCGT**GCC**CGGGGATTTGGTGC	68
	S53A Rev	GGCTTCG**GGC**GTGCCCGGGGATTT	69
	S82A For	GGCTGG**GCT**CTGGGCGGGCTGGTGGCAAGC	75
	S82A Rev	GCCCAGAGCCC**AGC**CTAACCAAATGGCTTTATCAGGTGCC	72
	D207A For	TATCTC**GCC**GGTCTGGTGCCGCGCAAAGTG	70
	D207A Rev	CCGTTTTTGCGATTGTAT**GGC**TATCTCGCCGGTCTG	68
	H235A For	GCGGCC**GCT**GCGCCATTTATTTCGCAT	68
	H235A Rev	TGGCGC**AGC**GGCCGCTTTGGCGAAGATATATGA	71
BsQueC	**K163A For**	CTCAAT**GCG**GCAGAAACGTGGAAGCTT	64
	K163A Rev	TTCTGC**CGC**ATTGAGCCACATGAGAGG	65
	Y187A For	ACATGT**GCT**AACGGCATCATCGCAGAC	64
	Y187A Rev	GCCGTT**AGC**ACATGTCAGCGTGTTGTT	64
	R204A For	CACCTT**GCC**TCAAAAGGGTATGAAGAA	59
	R204A Rev	TTTTGA**GGC**AAGGTGGCATGC	59
SrToyM	**K176A For**	GGAGC**GCG**ACGGACATCG	59
	K176A Rev	GTCCGT**CGC**GCTCCAGTGGAT	58
	R213A For	TACGAA**GCC**CGTGAGGCATTC	60
	R213A Rev	CTCACG**GGC**TTCGTAACACGT	60

Table 5.5: Sequences of the site-directed mutagenesis primers used in this work. Modified sequences are highlighted in bold

PCR reactions were prepared on ice and consisted of: 150 ng template DNA, 10 µM forward primer (1.5 µL), 10 µM reverse primer (1.5 µL), 10 mM dNTP mix (1.0 µL), 5x Phusion HF reaction buffer (10 µL), Pfu DNA polymerase enzyme (1.0 µL) and deionised water to a final volume of 50 µL. Primer sequences showing the restriction sites and PCR conditions are showed in Table 5.6.

Step	Temperature (°C)	Time (min)	Number of Cycles
Initial denaturation	98	2	1
Denaturation	98	1	30
Annealing	T_m - 5	0.5	
Extension	72	2	
Final Extension	72	5	1

Table 5.6: PCR reactions conditions

15 µL of the reaction was analysed by gel electrophoresis to check for correct amplification. Amplified PCR products were mixed with CutSmart buffer (4 µL) and Dpn1 enzyme (1.5 µL). The reaction was incubated at 37°C overnight. 5 µL of the mix were used to transform C2987 cells on LB Agar.

5.2.3 EcBioH initial cloning

5.2.3.1) Amplification and purification of ECBioH gene

Gene amplification was done by polymerase chain reaction (PCR) using *E.Coli* genomic DNA as a template and a pair of forward and reverse primers. These primers contain restriction sites to allow ligation and digestion of the gene on the plasmid (Nde1 for the Forward, Xho1 for Reverse). Primer sequences showing the restriction sites and PCR conditions are showed on Table 5.7.

Step	Temperature (°C)	Time (min)	Number of Cycles
Initial denaturation	98	2	1
Denaturation	98	1	30
Annealing	55	0.5	
Extension	72	1	
Final Extension	72	5	1
Forward Primer	CGATCATATGAATAACATCTGGTGGCAGACCAAAGGT		
Reverse Primer	CGATCTCGAGCTACACCCTCTGCTTCAACGCCAC		

Table 5.7: PCR reactions conditions and primers sequences. Restriction sites are showed in red.

The amplified gene was then A-tailed to add a single deoxyadenosine to the 3' end, in order to allow ligation into a pGEM-T vector. This was done by incubating the PCR reaction mixture at 72°C for

25 mins in presence of PureTaq beads, containing Taq DNA polymerase, nucleotides and stabilizers. The amplified and A-tailed gene was then purified on a 1% agarose gel containing 0.01% Gel Red, revealed by UV light and extracted from the gel using a ThermoScientific gel extraction kit.

5.2.3.2) pGEM ligation and analysis

To ease the analysis of the gene, it was first cloned onto a pGEM-T vector because it is an efficient way to check if the amplified gene was correctly ligated onto the vector, as this vector is compatible with blue-white screen using X-Gal.

First, the gene was ligated in the pGEM-T vector. The ligation reaction consisted of 150 ng of the amplified gene, 1 µL pGEM-T vector, 1 µL pGEM-T reaction buffer and 1 µL T4 ligase. The reaction was incubated overnight at 4°C.

The ligation mixture (4 µL) was then transformed into C2987 highly competent cells (25 µL) in LB Agar / X-Gal plates supplemented with Ampicillin (100 µg/mL). Such plates were used to do a blue-white screening. White colonies should indeed contain the insert while blue colonies do not. Hence, five white colonies were picked on the following day and inoculated into five aliquots of LB broth (10 mL) supplemented with Ampicillin (100 µg/mL). These aliquots were shaken for 16 hours at 37C and centrifuged at 3500 rpm for 10 mins separately. The supernatant was discarded, while the DNA of the cell pellets was extracted with a ThermoScientific GeneJET miniprep kit following the manufacturer procedure. The DNA of each colonies was eluted separately into 50 µL elution buffer.

To further verify the presence of the insert, analytical digest of the five products was undergone. The analytical digest consisted of 1 µL DNA from the previous step, 1 µL of each restriction enzymes (Nde1 and Xho1), 1 µL CutSmart buffer and 6 µL water. This was let at 37°C for 3 hours and ran on an 1% agarose gel supplemented with 0.01% GelRed and revealed under UV light. DNA products in which the insert was observed were then sequenced, others were discarded.

5.2.3.3) Vector transfer of the insert: from pGEM-T vector to pET28a vector

When the sequence was found to be the correct one, preparatory digest was done to cut off the insert from the pGEM-T vector as well as to open the circular pET28a vector in which the insert will subsequently be ligated. The preparatory digest reaction consisted of 50 µL DNA (insert + pGEM-T vector OR pET28a vector), 6 µL CutSmart buffer and 1 µL of each restriction enzymes (Nde1 and Xho1). The reaction was allowed to occur for 16 hours at 37°C. The opened pET28a vector could be used without any further purification but the insert had to be separated from the pGEM-T vector and was therefore ran on a 1% agarose gel supplemented with 0.01% GelRed. The insert, revealed under UV

light, and extracted from the gel using a ThermoScientific gel extraction kit, following the manufacturer procedure.

The insert was subsequently ligated into the pET28a vector. This ligation reaction was proceeded as previously described for the pGEM-T vector. The ligation mixture (4 µL) was then transformed into C2987 highly competent cells (25 µL) in LB Agar plates supplemented with Kanamycin (30 µg/mL). Twelve colonies were picked on the following day and inoculated into twelve aliquots of LB broth (10 mL) supplemented with Kanamycin (30 µg/mL). These aliquots were shaken for 16 hours at 37C and centrifuged at 3500 rpm for 10 mins separately. The supernatant was discarded, while the DNA of the cell pellets was extracted with a ThermoScientific GeneJET miniprep kit following the manufacturer procedure. The DNA of each colonies was eluted separately into 50 µL elution buffer.

To verify the presence of the insert, analytical digest of the products was undergone. The analytical digest consisted of 1 µL DNA from the previous step, 1 µL of each restriction enzymes (Nde1 and Xho1), 1 µL CutSmart buffer and 6 µL water. This was let at 37°C for 3 hours and ran on an 1% agarose gel supplemented with 0.01% GelRed and revealed under UV light. DNA products in which the insert was observed were then used to express BioH, others were discarded.

5.3 Protein production

5.3.1 Transformation

Plasmid DNA (2.5 µL) was added to an aliquot of competent *E. coli* cells (10 or 25 µL) and incubated on ice for 30 mins. The cells were subsequently heat shocked at 42°C for 45 seconds. SOC (120 µL) was added and the cells were incubated at 37°C with shaking (200 rpm) for 1 h. The mix was then transferred on an LB Agar plate supplemented with the appropriate antibiotic (ampicillin: 100 mg/L or kanamycin: 30 mg/L) and the cells were grown for 16 hours at 37°C.

5.3.2 Expression tests

A single colony of BL21 (DE3) *E. coli* cells containing the expression plasmid was used to inoculate 10 mL LB + appropriate antibiotic and shaken at 200 rpm for 16 hours at 37°C. The optical density at 600 nm (OD_{600}) was then recorded and the pre-culture was diluted in 10 mL fresh LB + appropriate antibiotic to an OD_{600} of 0.15. The new culture was allowed to grow to an OD_{600} of 0.6-0.9 at 37°C. IPTG was added to a final concentration of 0.1 mM and the cultures were incubated further at 16°C and 30°C for 5 hours and overnight. The cell pellets were harvested by centrifugation (3500 rpm for 15 mins) and resuspended in phosphate buffer saline (PBS) (1 mL). Protein expression was analysed after cell lysis by sonication (15 cycles, 10 secs on/10 secs off) and separation of cell pellets and supernatant

by centrifugation (13000 rpm for 10 mins) by denaturing 12% SDS-PAGE gel run at 200 V. Optimal conditions were use for large scale expression.

5.3.3 Large scale expression

A single colony of BL21 (DE3) *E. coli* cells containing the expression plasmid was used to inoculate 250 mL LB + appropriate antibiotic and shaken at 200 rpm for 16 hours at 37°C. The optical density at 600 nm (OD_{600}) was then recorded, and the pre-culture was diluted in 1 L fresh LB + appropriate antibiotic to an OD_{600} of 0.15. The new culture was allowed to grow to an OD_{600} of 0.6-0.9 at 37°C. After which IPTG (to a final concentration of **a** mM) and $ZnSO_2$ (to a final concentration of **b** mM) were added to induce protein expression. The cultures were shaken further for **c** hours at **d**°C. Bolded parameters are shown in Table 5.9.

	IPTG concentration (mM)	$ZnSO_2$ concentration (mM)	Expression duration (h)	Expression temperature (°C)
Parameter	**a**	**b**	**c**	**D**
EcBioH WT and mutants, HiBioG WT	0.5	0	18	18
BsQueC WT and mutants	0.1	0.1	5	30
SrToyM WT and mutants, BsQueF WT	0.1	0.1	18	16

Table 5.9: Parameters used for large scale expression for each protein

5.3.4 Expression of SeMet-SrToyM

A single colony of B834 (DE3) *E. coli* cells containing the SrToyM expression plasmid was inoculated in 500 mL LB + kanamycin (30 mg/L) and shaken at 200 rpm for 16 hours at 37°C. The cells were then harvested by centrifugation (3000 rpm, 10 minutes). The supernatant was decontaminated with Virkon and discarded. The pellet was re-suspended in sterilised SelenoMet™ base medium (30 mL) and centrifuged (3000 rpm, 10 minutes). This was repeated three times to ensure no traces of LB medium remained. The supernatant was decontaminated with Virkon and discarded in each case. The pellet was resuspended in SelenoMet™ base medium (30 mL). On the other hand, 2 L sterilised SelenoMet™ base medium supplemented by SelenoMet™ nutrient mix (50 mL) and kanamycin (30 mg/L) were prepared. To these cultures were added the cell suspension to a final OD_{600} of 0.05. These cultures were grown at 37°C with shaking (200 rpm) for 20 min. 10 mL of each culture were taken as controls and selenomethionine (30 mg/L) was added. The cells were then grown at 37°C with shaking

(200 rpm) to an OD_{600} of 0.6. Protein expression was then induced by addition of IPTG (to a final concentration of 0.1 mM) and $ZnSO_2$ (to a final concentration of 0.1 mM). The cultures were shaken further for 18 hours at 16°C.

5.3.5 Cell harvesting

The bacterial cells were harvested by centrifugation in a ThermoScientific Multifuge centrifuge at 7000 rpm for 7 min at 4°C. The supernatant was disinfected with Virkon and discarded, and the cell pellet was resuspended in a minimum amount of PBS and further centrifuged at 3500 rpm for 15 mins at 4°C. The supernatant was discarded, and the cell pellets, containing the protein of interest was stored at -20°C.

5.4 Protein purification

5.4.1 Cell lysis

Cells were resuspended in buffer A/C/F and lysed by sonication for 15 cycles (30 seconds on, 30 seconds off). DNase (0.2 mg) was added and the lysed cell suspension was centrifuged (12000 rpm, 45 mins). The supernatant was filtered with a 0.45 µm filter. Protein was purified from the filtered cell-free extract.

5.4.2 IMAC purification

All purification steps were carried out at 4°C or on ice. Histidine-tagged proteins were purified by Nickel immobilised metal affinity chromatography (IMAC) using 1 mL or 5 mL HisTrap HP columns attached to an AKTA Start. The column was washed with 10 column volumes (CVs) of Buffer **a** and the cell-free extract was loaded at **x** mL/min. The column was washed with Buffer **a** for 7 CVs and the protein of interest was eluted with a gradient between Buffer **a** and Buffer **b** for 10 CVs at a flow rate of **x** mL/min, monitoring at 280 nm. The fractions that contained the protein were then pooled and concentrated to 5 mL using a Vivaspin 20 (GE Healthcare, MW cut-off = 10 kDa) before SEC. Bolded parameters are shown in Table 5.10.

	Wash buffer (from Table 5.4)	Elution buffer (from Table 5.4)	Flow rate (mL/min)
Parameter	**a**	**b**	**x**
EcBioH WT and mutants, HiBioG WT	C	D	
BsQueC WT and mutants, BsQueF WT, SrToyM WT and mutants	F	G	
1 mL HisTrap HP column			1.0
5 mL HisTrap HP column			5.0

Table 5.10: Parameters used for the IMAC purification

5.4.3 Anion exchange purification of untagged EcBioH

Untagged EcBioH was purified by anion exchange chromatography using a HiPrep Q FF 16/10 column attached to an AKTA Explorer. The column was washed with 5 CVs of Buffer C and the cell-free extract was loaded at 1.6 mL/min. The column was washed with Buffer C for 5 CVs and the protein of interest was eluted with a gradient between Buffer A and Buffer B for 10 CVs at a flow rate of 1.6 mL/min, monitoring at 280 nm. The fractions that contained the protein were then pooled and concentrated to 5 mL using a Vivaspin 20 (GE Healthcare, MW cut-off = 10 kDa) before SEC.

5.4.4 Size-exclusion chromatography

The concentrated protein (5 mL) was then loaded onto a pre-equilibrated (1 CV, buffer **a**) HiPrep 16/600 Superdex **b.** The recombinant enzyme was eluted at a flow rate of **c** mL/min, monitoring at 280 nm. The purity of recombinant proteins was analysed by 12% SDS-PAGE. Fractions containing the protein were combined. If the protein was to be used for crystallography, it was concentrated to 10 mg/mL using a Vivaspin 20 (GE Healthcare, MW cut-off = 10 kDa) and kept on ice at 4°C. Otherwise, the enzyme was concentrated/diluted to 4 mg/mL, flash-frozen in liquid nitrogen and stored at -80°C. Bolded parameters are shown in Table 5.11.

	Buffer (from Table 5.4)	Column	Flow rate (mL/min)
Parameter	**a**	**b**	**c**
EcBioH WT and mutants, HiBioG WT	E	S-75	0.6
BsQueC WT and mutants, BsQueF WT, SrToyM WT and mutants	H	S-200	1

Table 5.10: Parameters used for the size-exclusion chromatography

5.5 Protein characterisation

5.5.1 Determination of enzyme concentration using Bradford assay

In all cases, protein concentration in a sample was determined using the Bradford assay. Coomasie brilliant blue G-250 dye (shown in Figure 5.1) undergoes a colour change from brown to blue upon interacting with the basic residues side chains (Lys, Arg, His) contained in an enzyme.

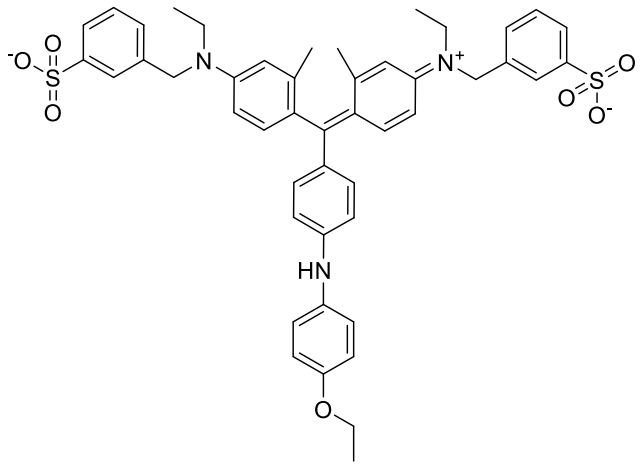


Figure 5.1: Structure of the Coomasie brilliant blue G-250 dye

Five standard solutions of BSA (50 µL) at various concentrations (0 mg/mL, 0.25 mg/mL, 0.5 mg/mL, 1 mg/mL and 1.4 mg/mL) were prepared and mixed with Bradford reagent (1.5 mL). The absorbance of these solutions at 595 nm was recorded and used to plot a calibration curve. 50 µL of a solution containing the enzyme with unknown concentration was then diluted appropriately for its absorbance to fit within the calibration curve range, mixed with Bradford reagent (1.5 mL) and its absorbance at 595 nm was recorded.

5.5.2 SDS-PAGE

A routine gel consisted of a 15% acrylamide running gel topped by a 4% acrylamide stacking gel. The running gel was made by mixing deionised water (5.7 mL), 40% acrylamide (6 mL), 1.5 M Tris (pH = 8.8, 4 mL), 10 w/v SDS (150 µL), 50 mg/mL ammonium persulfate (APS) (350 µL) and tetramethylethylenediamine (TEMED) (20 µL). The running gel was made by mixing deionised water (2.9 mL), 40% acrylamide (750 µL), 0.5 M Tris (pH = 6.8, 1.25 mL), 10 w/v SDS (50 µL), 50 mg/mL APS (100 µL) and TEMED (5 µL). Protein samples (10 µL) were prepared by adding SDS loading buffer (5 µL) containing: 0.5 M Tris (pH = 6.8, 2.5 mL), glycerol (2 mL), 10 w/v SDS (4 mL), β-mercaptoethanol (1 mL) and 0.1% w/v bromophenol blue (500 µL). The samples were then boiled for 5 minutes and loaded onto gels alongside Low molecular weight marker (GE Healthcare) and were run for 50 minutes at a constant voltage of 200 V. Gels were revealed by immersion in Instant Blue Coomasie stain (Expedeon) for 30 minutes at room temperature (RT) with gentle shaking and then washed with water before revelation under light.

5.5.3 Protein mass spectrometry

The protein was diluted to a concentration of 10 µM using deionised water and centrifuged for 5 minutes at 13000 rpm. The supernatant (5 µL) was injected on a Synapt G2-Si Q-TOF (Waters) instrument, that comprised a Phenomenex C4 2.6 µm liquid chromatography (LC) column coupled with an ESI source. The LC gradient ran from 5% to 95% acetonitrile in water in 12 minutes. Both solvents were supplemented by 0.1% formic acid. The MS source was set at 120°C and the sampling cone voltage was 54 V. Protein spectra was smoothened, centered and subtracted using MassLynx V4.1.

5.6 Organic synthesis

5.6.1 Model MBH reaction

To a stirred solution of 4-NBA (93 mg, 0.62 mmol, 1eq) in tetrahydrofuran (THF) (1 mL) were added successively DABCO (11 mg, 0.093 mmol, 0.15 eq) and MVK (130 mg, 1.85 mmol, 3eq). The reaction mixture was stirred for 20 hours at RT. The product was extracted with dichloromethane (DCM) (3*15 mL). The combined organics were dried with magnesium sulfate and concentrated in vacuuo to afford

the crude product. Flash column chromatography was then carried. Mobile phase was a gradient 0-40% ethyl acetate (EtOAc) in Petroleum ether (30-40°C). The desired product was obtained (46 mg, 0.21 mmol, 34%). Nuclear magnetic resonance (NMR) and ESI-MS analyses showed that the product has >95% purity.

^{1}H NMR (500 MHz, dCDCl$_3$): δ = 8.21-8.23 (2H, d, CH-1,3), 7.59 (2H, d, CH-4,6), 6.28 (1H, s, CH-12), 5.71 (1H, s, OH-10), 2.27 (1H, s, CH-8), 2.38 (3H, s, CH$_3$-13).

MS (ESI, [M - H$^+$]): *m/z* = 220.0949 (Theoretical *m/z* ($C_{11}H_{11}O_3N$) = 220.067)

5.6.2 Organic synthesis of Chapter 4 compounds

5.6.2.1. Synthetic route to CDG and PreQ$_0$

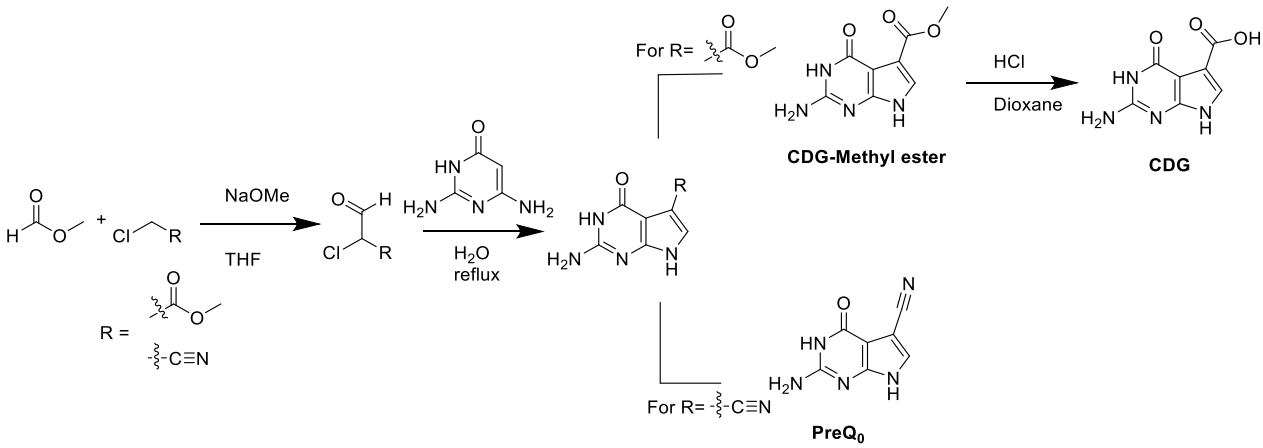
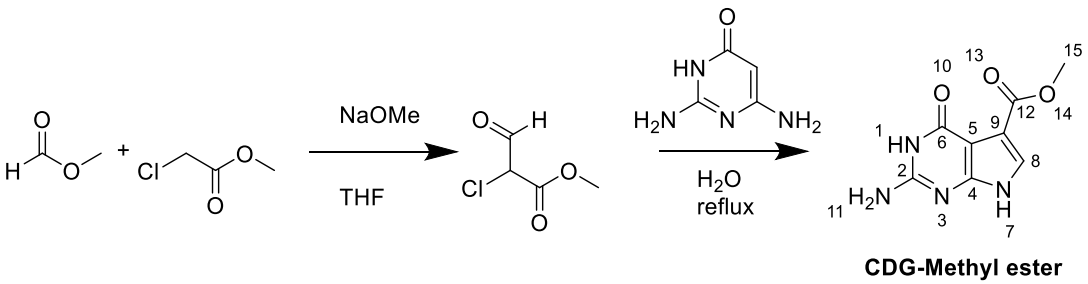


Figure 5.1: CDG and PreQ$_0$ synthetic schemes.

5.6.2.2 One-pot synthesis of CDG-methyl ester



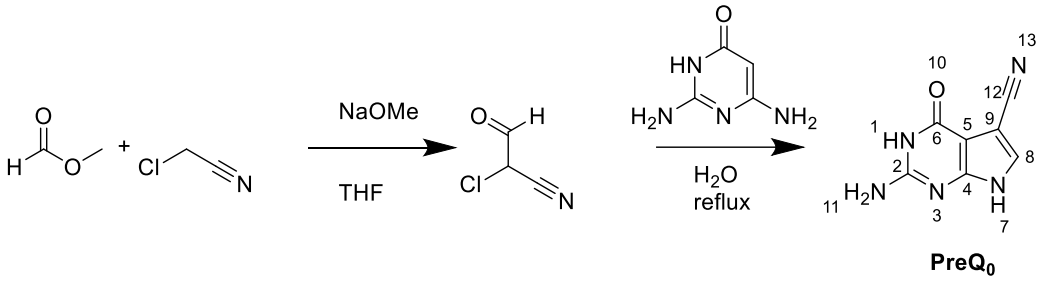
To methyl formate (3.20 g, 58 mmol, 5.8 eq.) in THF (25 mL) cooled to 0°C were subsequently added sodium methoxide (0.72 g, 13 mmol, 1.3 eq.) and methyl chloroacetate (1.09 g, 10 mmol, 1.0 eq). The reaction was stirred overnight at RT. Deionised water (25 mL) was added to stop the reaction and quench excess sodium methoxide. Excess methyl formate and THF were removed under reduced pressure and 2,4-diamino-6-hydroxypyrimidine (1.20 g, 10 mmol, 1.0 eq.) was subsequently added

and the mixture was refluxed for 1 hour. The product precipitated upon cooling and was recovered by filtration. After THF washings, the product was obtained (0.92 g, 3.7 mmol, 47% yield) with a purity exceeding 95% (^{1}H NMR, ^{13}NMR)

^{1}H NMR (600 Hz, dDMSO): δ = 11.63 (1H, s, NH-1), 10.35 (1H, br s, NH-7), 7.35 (1H, d, *J* = 3Hz, CH-8), 6.49 (2H, br s, NH_2-11), 2.80 (3H, s, CH_3-15)

^{13}C NMR (600 Hz, dDMSO): δ = 163.80 (C-12), 160.30 (C-6), 153.66 (C-2), 153.34 (C-4), 125.40 (C-8), 110.27 (C-9), 97.85 (C-5), 52.15 (C-15)

5.6.2.3 One-pot synthesis of $PreQ_0$



$PreQ_0$

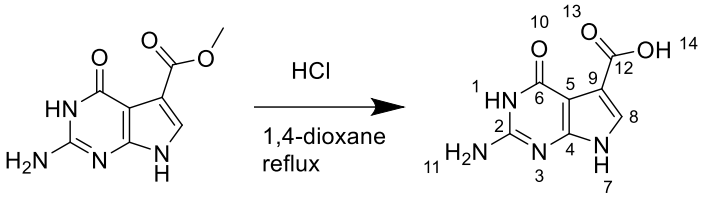
To methyl formate (3.20 g, 58 mmol, 5.8 eq.) in THF (25 mL) cooled to 0°C were subsequently added sodium methoxide (0.72 g, 13 mmol, 1.3 eq.) and chloroacetonitrile (0.76 g, 10 mmol, 1.0 eq). The reaction was stirred overnight at RT. Deionised water (25 mL) was added to stop the reaction and quench excess sodium methoxide. Excess methyl formate and THF were removed under reduced pressure and 2,4-diamino-6-hydroxypyrimidine (1.20 g, 10 mmol, 1.0 eq.) was subsequently added and the mixture was refluxed for 1 hour. The product precipitated upon cooling and was recovered by filtration. After THF washings, the product was obtained (1.12 g, 6.4 mmol, 64% yield) with a purity exceeding 95% (^{1}H NMR, ^{13}NMR) (compared with literature data[1]).

^{1}H NMR (600 Hz, dDMSO): δ = 11.63 (1H, s, NH-1), 10.35 (1H, br s, NH-7), 7.35 (1H, d, *J* = 3Hz, CH-8), 6.49 (2H, br s, NH_2-11)

^{13}C NMR (600 Hz, dDMSO): δ = 158.09 (C-12), 154.26 (C-6), 152.16 (C-2), 128.40 (C-4), 116.37 (C-8), 99.19 (C-9), 86.00 (C-5)

MS (ESI, [M + H^+]): *m/z* = 176.0558 (Theoretical *m/z* ($C_7H_5N_5O$) = 176.0572)

5.6.2.4 Synthesis of CDG



To a suspension of CDG-methyl ester (0.92 g, 3.7 mmol, 1.0 eq.) in 1,4-dioxane (10 mL) was added HCl (37%, 10 mL, excess). The mixture was refluxed for 5 hours. The precipitate formed upon cooling was filtered, washed with copious amounts of THF and dried. The brown solid (0.67 g, 2.8 mmol, 81%) was confirmed to be the product (^{1}H NMR, ^{13}C NMR and ESI-MS) at purity > 95% (compared with literature data[1]).

^{1}H NMR (600 Hz, dDMSO): δ = 11.96 (1H, s, NH-1), 11.63 (1H, br s, OH-14), 11.13 (1H, br s, NH-7), 7.43 (1H, d, *J* = 3Hz, CH-8), 6.44 (2H, br s, NH_2-11)

^{13}C NMR (600 Hz, dDMSO): δ = 163.48 (C-12), 161.86 (C-6), 153.55 (C-2), 152.74 (C-4), 126.30 (C-8), 110.98 (C-9), 96.82 (C-5)

MS (ESI, [M + H^{+}]): *m/z* = 195.0509 (Theoretical *m/z* ($C_7H_6O_3N_4$) = 195.0518)

5.7 Protein reactions for Chapter 3

5.7.1 Biocatalysed MBH reaction optimised protocol

To a solution of EcBioH (WT or mutant) in Buffer E (4 mg/mL) were subsequently added 4-NBA (8 mM, from 50x stock in acetonitrile) and MVK (64 mM, from 50x stock in acetonitrile). The reaction mixture was shaken at 280 rpm at 30°C for 96h.

5.7.2 Biocatalysed MBH reaction HPLC assay

To a sample of the reaction mixture (15 µL) was added 100% trichloroacetic acid (TCA, 3 µL). The sample was centrifuged at 13000 rpm for 2 mins. To the supernatant was added 20% MeCN in water (100 µL). 10 µL of the sample was injected on a Phenomenex Luna 5u C18 HPLC column and eluted with 50% MeCN + 0.1% TFA in water + 0.1% TFA at a flow rate of 1 mL/min for 8 minutes. 4-NBA and the MBH product were detected by a UV detector at 254 nm, while MVK does not absorb at this wavelength. Concentrations of the analytes were determined by comparing the area under the curve (AUC) with a calibration curve.

5.7.3 Biocatalysed MBH reaction chiral HPLC assay

A sample of the reaction mixture (200 µL) was extracted with ethyl acetate (EtOAc, 3*200 µL). The solvent was removed under reduced pressure and the solid was redissolved in isopropanol (100 µL). 10 µL of the sample was injected on a HiChrom ChiralPak IC column and eluted with 8% isopropanol in hexane at a flow rate of 1 mL/min for 50 minutes. 4-NBA and the MBH product were detected by a UV detector at 254 nm, while MVK does not absorb at this wavelength.

5.7.4 pNPA esterase assay

The ECBioH and HiBioG esterase activity were assayed by monitoring the hydrolysis of pNPA (ξ_{410} = 18 L/mmol/cm). The final volume of the reactions was 250 µL and contained 0-4 mM pNPA from a 10 mM stock in isopropanol, 3.4 mg/mL Triton X-100 and 50 mM Tris (pH = 8.0). These reagents were preincubated and mixed at 37°C and the reaction was blanked. The reaction was initiated by the addition of the enzyme (67 µM). The increase of absorbance at 410 mM from the enzymatic conversion of pNPA into pNP was monitored over 1h on a BioTek Synergy HT plate reader with Costar 96-well plate. The data was analysed using the Michaelis-Menten model and a nonlinear regression fit on OriginLab 2016 was used to find the values of K_M and k_{cat}.

5.7.5 MS assay

5.7.5.1 PMSF

To EcBioH (1 mg/mL) was added PMSF (1 mM). The mixture was stirred for one hour at 37°C after which cyclohexenone (1 mM) was added. The mix was then stirred for 90 minutes and analysed by ESI-MS as developed in Part 5.5.3. Controls, where either PMSF or cyclohexenone was not added, were also prepared and assayed.

5.7.5.2 IAA

To EcBioH (1 mg/mL) was added IAA (final concentration: 1 mM). The mixture was stirred for one hour at 37°C in the dark after which cyclohexenone (1 mM) was added. The mix was then stirred for 90 minutes and analysed by ESI-MS as developed in Part 5.5.3. Controls, where either IAA or cyclohexenone was not added, were also prepared and assayed. The IAA-only control was also assayed for MBH activity following the method described in section 5.7.1.

5.7.6 Buffer exchange

The enzymes and proteins used in Part 2.8.6 as well as BioH S82A D207A were buffered-transferred into Buffer E by diafiltration. The protein sample in the original buffer was loaded onto a

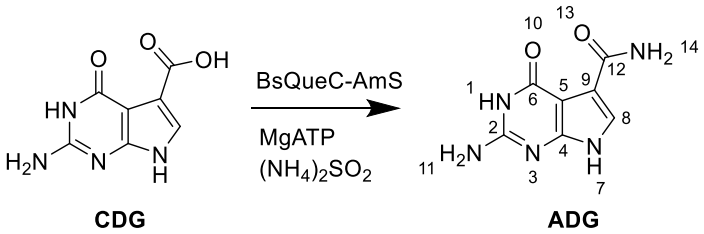
Vivaspin 20 (GE Healthcare, MW cut-off = 10 kDa) spin column. The sample was concentrated to a volume of 1 mL after which 15 mL of Buffer E were added. The sample was re-concentrated to a volume of 1 mL, and this operation was repeated twice. Finally, 10 mL of Buffer E were added and the sample was concentrated to an enzyme concentration of 4 mg/mL and assayed for the MBH reaction following the protocol described in section 5.7.1.

5.8 Protein reactions for Chapter 4

5.8.1 NS reaction optimised protocol

To a solution of NS in Buffer H (2 mg/mL) were added magnesium chloride (50 mM, from 10x stock in deionised water), ammonium sulfate (50 mM, from 10x stock in deionised water), ATP (1 mM, from 100x stock in deionised water) and CDG (0.5 mM, from 200x stock in DMSO). The mix was then stirred at 280 rpm for 5 hours at 37°C.

5.8.2 Large scale biosynthesis of ADG using BsQueC-eAmS



CDG **ADG**

To a 350 mL solution of BsQueC-eAmS in Buffer H (0.25 mg/mL) were subsequently added magnesium chloride (50 mM, excess, from 10x stock in deionised water), ammonium sulfate (50 mM, excess, from 10x stock in deionised water), ATP (386 mg, 0.7 mmol, 2 mM, 2 eq., from 50x stock in water) and CDG (67 mg, 0.35 mmol, 1 mM, 1 eq., from 100x stock in DMSO). The reaction was shaken at 280 rpm for 2 hours at 37°C and filtered. The recovered solid was redissolved in DMSO (20 mL) and filtered on a centrifugal spin column (Vivaspin 20 (GE Healthcare, MW cut-off = 10 kDa). The filtrate was kept and the DMSO was removed under high vacuum (5 days). The product was finally washed with copious amounts of THF and recovered by filtration. The brown solid (9 mg, 0.05 mmol, 14%) was confirmed to be ADG (^{1}H NMR, ^{13}C NMR and ESI-MS) at purity > 95% (compared with literature data[1]).

^{1}H NMR (600 Hz, dDMSO): δ = 11.57 (1H, s, NH-1), 10.68 (1H, br s, NH-7), 9.55 (1H, s, NH-14), 7.21 (1H, d, CH-8), 6.87 (1H, s, NH-14), 6.37 (2H, br s, NH_2-11)

^{13}C NMR (600 Hz, dDMSO): δ = 164.75 (C-6), 160.76 (C-12), 153.20 (C-5), 153.06 (C-4), 123.43 (C-2), 115.08 (C-8), 96.47 (C-9)

5.8.3 HPLC assay

To samples of the reaction mixture (100 µL) was added 100% trichloroacetic acid (TCA, 5 µL). The samples were centrifuged at 13000 rpm for 2 mins. 10 µL of the supernatant was injected onto a Phenomenex Luna 5u C18 HPLC column. Samples were eluted with water + 0.1% TFA for 5 minutes followed by a 17 minutes gradient to 20% MeCN + 0.1% TFA. The column was then washed for 1 minute with 50% MeCN + 0.1% TFA/50% water + 0.1% TFA before reequilibrating the column with 100% water + 0.1% for 5 minutes. The analytes were detected by a UV detector at 254 nm.

5.8.4 BsQueF reductase assay

BsQueF nitrile reductase activity was assayed by monitoring the hydrolysis of oxidation of NADPH into $NADP^+$. The final volume of the reactions was 250 µL and contained NADPH (1 mM) and $PreQ_0$ (0-500µM). These reagents were preincubated and mixed at 37°C, and the reaction was blanked. The reaction was initiated by the addition of BsQueF (1 mg/mL). The decrease of absorbance at 340 mM from the enzymatic conversion of NADPH into $NADP^+$ was monitored over 1h on a BioTek Synergy HT plate reader with Costar 96-well plate. The data was analysed using the Michaelis-Menten model, and a nonlinear regression fit on OriginLab 2016 was used to find the values of K_M and k_{cat}.

5.8.5 BsQueC-eAmS kinetics study

The activity of BsQueC-eAmS amide was measured by monitoring the conversion of CDG to ADG by reverse-phase HPLC. 1.2 mL reaction mixtures containing BsQueC-eAmS in Buffer H (1 mg/mL), magnesium chloride (50 mM, from 10x stock in deionised water), ammonium sulfate (50 mM, from 10x stock in deionised water), ATP (0-2 mM, from 100 mM stock in deionised water) and CDG (0-1 mM, from 100 mM stock in DMSO) were prepared. These mixtures were then shaken at 280 rpm at 37°C. At various time points (0.5 min, 1 min, 2 min, 5 min, 15 min, 30 min, 60 min, 150 min, 300 min) samples (100 µL) of the reaction were taken, quenched with TCA (5 µL) and analysed by HPLC following the procedure presented in section 5.8.3. The concentration of formed ADG at each timepoint was found and the data was analysed using the Michaelis-Menten model. A nonlinear regression fit on OriginLab 2016 was used to find the values of K_M and k_{cat}.

5.8.6 X-ray crystallography

5.8.6.1 Robot screen

After SEC, BsQueC and BsQueC-eAmS was concentrated to 7.5 mg/mL and SrToyM was concentrated to 10 mg/mL using a centrifugal spin column (Vivaspin 20 (GE Healthcare, MW cut-off =

10 kDa). Ligands were added at this stage as required and the protein sample was centrifuged for 10 minutes at 14000 rpm to remove any aggregates. The supernatant was taken to set up high-throughput screens using an ARI Robotics Gryphon robot and commercial screens (Hampton Research PEG ION, Molecular Dimensions JCSG+ and Molecular Dimensions Structures 1+2). Crystals were formed by vapour diffusion following the sitting-drop method. The screens were set up on a 96 wells MRC plate with 100 nl protein solution mixed with 100 nl mother liquor and equilibrated against 200 nl mother liquor. Crystals were then allowed to grow for seven days at 23°C. Conditions yielding crystals were further optimised

5.8.6.2 Optimisation protocol

After SEC, the enzyme to crystallise was concentrated to the corresponding concentration (7.5 mg/mL for BsQueC and BsQueC-eAmS; 10 mg/mL for SrToyM) using a centrifugal spin column (Vivaspin 20 (GE Healthcare, MW cut-off = 10 kDa). Ligands were then added as required and the protein sample was centrifuged for 10 minutes at 14000 rpm to remove any aggregates. Crystals were formed by vapour diffusion following the hanging-drop method. The optimisation was carried on a 24-well plate. 1 mL of mother liquor was added to each well and each crystallisation drops was prepared by mixing various ratios (1:2, 1:1, 2:1) of protein solution and mother liquor. Crystals were grown for 7 days at 23°C. Suitable crystals were cryoprotected using the same mother liquor supplemented by 20% w/v glycerol and flash-frozen by immersion in liquid nitrogen. Datasets were collected on beamline I03 at the Diamond Light Source (Didcot, UK) at 100 K using a Pilatus 6 M detector, in collaboration with Dr Jon Marles-Wright (University of Newcastle).

5.8.6.3 ADP-bound BsQueC structure

BsQueC crystals trials were conducted in the presence of 2.5 mM ADP and 12.5 mM magnesium chloride. The crystals which were grown in 0.2 M magnesium chloride hexahydrate, 0.1 M Tris (pH 7), 7% w/v PEG 8000, 2.5 mM ADP and 12.5 mM magnesium chloride has provided the best quality diffraction, to a resolution of 2.1 Å. The structure of BsQueC + ADP was determined by molecular replacement using BsQueC structure with PDB code 3BL5. Creation and refinement of the model were carried by Dr Jon Marles-Wright (University of Newcastle). ADP was fitted using Coot. Crystallographic figures were generated with PyMOL (Schrodinger LLC). Data collection and refinement statistics are shown in the Appendix.

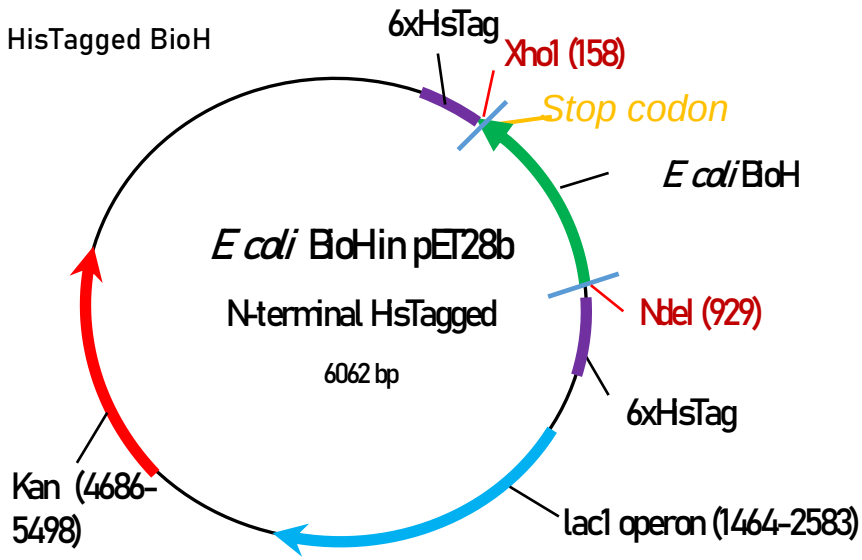
References:

[1] A. Gangjee, A. Vidwans, 2001, *J.Med.Chem.,* 44, pp 1993-2003

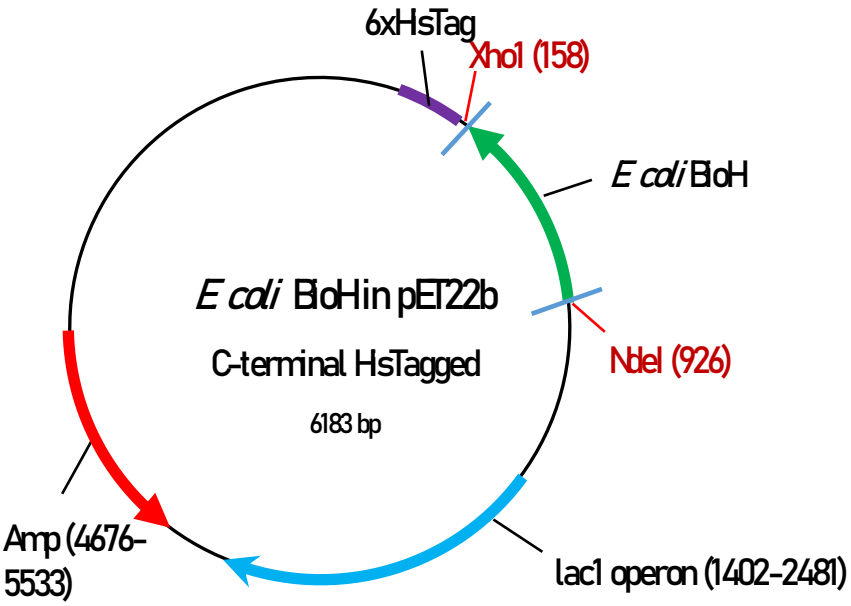
Appendix

Appendix 1: EcBioH vector maps

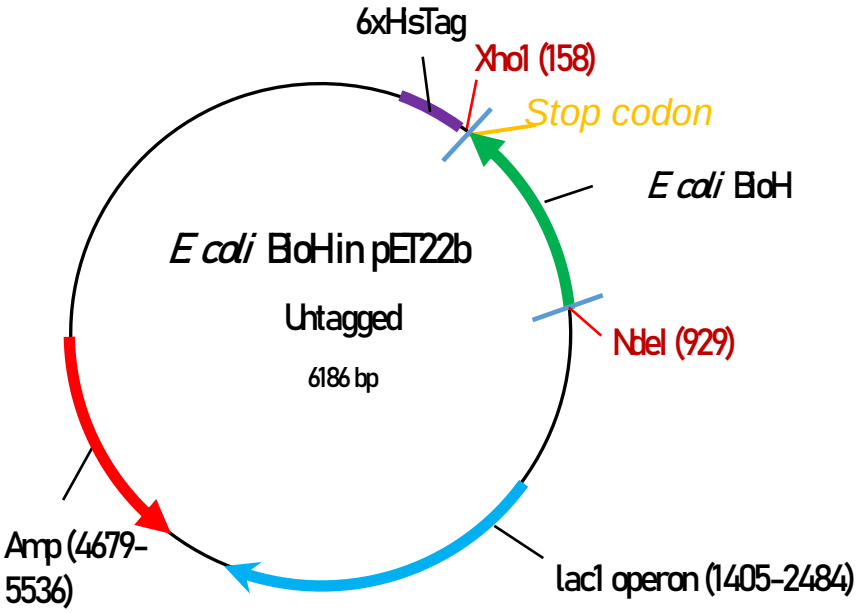
N-terminal HisTagged BioH



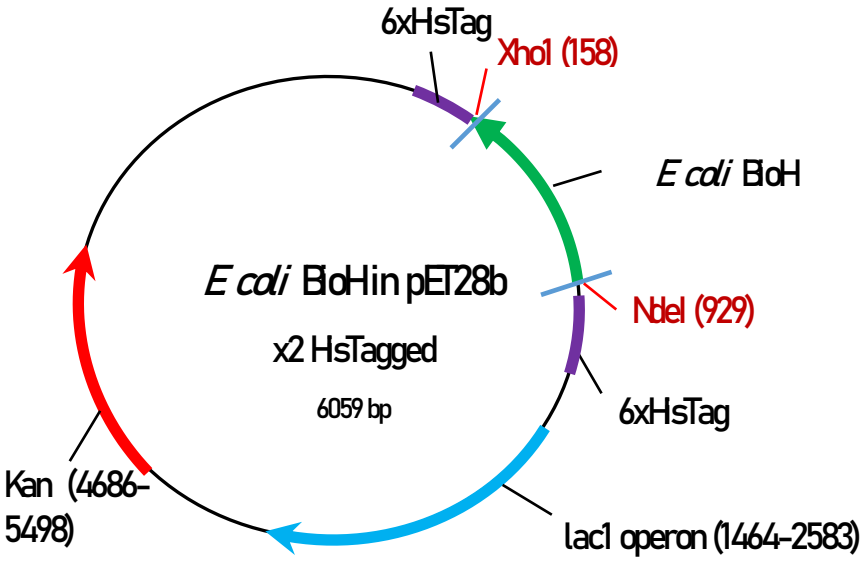
C-terminal HisTagged BioH



Untagged BioH



X2- HisTagged BioH



Appendix 2: *E. coli bioH* gene sequence

ATGAATAACATCTGGTGGCAGACCAAAGGTCAGGGGAATGTTCATCTTGTGCTGCTGCAC
GGATGGGGACTGAATGCCGAAGTGTGGCGTTGCATTGACGAGGAACTTAGCTCGCATTTT
ACGCTGCACCTTGTTGACCTGCCCGGCTTCGGGCGTAGCCGGGGATTTGGTGCGCTGTCA
CTTGCTGATATGGCCGAAGCCGTGCTGCAACAGGCACCTGATAAAGCCATTTGGTTAGGC
TGGAGTCTGGGCGGGCTGGTGGCAAGCCAGATTGCGTTAACCCATCCCGAGCGTGTTCAG
GCGCTGGTCACCGTGGCGTCGTCACCTTGTTTTAGTGCTCGTGACGAGTGGCCGGGGATA
AAACCGGACGTGCTGGCGGGATTTCAGCAGCAACTCAGTGATGATTTTCAGCGTACAGTG
GAGCGGTTCCTGGCGTTACAAACCATGGGGACTGAAACGGCGCGCCAGGATGCGCGGGC
GTTGAAGAAAACCGTTCTGGCGTTACCGATGCCGGAGGTTGACGTGCTTAATGGCGGGCTG
GAAATCCTGAAAACGGTCGATCTCCGTCAGCCGCTGCAAAACGTGTCCATGCCGTTTTTG
CGATTGTATGGCTATCTCGACGGTCTGGTGCCGCGCAAAGTGGTGCCGATGCTGGATAAA
CTTTGGCCTCACAGCGAATCATATATCTTCGCCAAAGCGGCCCATGCGCCATTTATTTCG
CATCCGGCCGAGTTTTGTCACCTGCTGGTGGCGTTGAAGCAGAGGGTGTAG

Appendix 3: EcBioH-N protein sequence (the hexahistidine tag is in bold)

MGSSHHHHHHSSGLVPRGSMNNIWWQTKGQGNVHLVLLHGWGLNAEVWRCIDEELSSHFTL
HLVDLPGFGRSRGFGALSLADMAEAVLQQAPDKAIWLGWSLGGLVASQIALTHPERVQALVTVA
SSPCFSARDEWPGIKPDVLAGFQQQLSDDFQRTVERFLALQTMGTETARQDARALKKTVLALP
MPEVDVLNGGLEILKTVDLRQPLQNVSMPFLRLYGYLDGLVPRKVVPMLDKLWPHSESYIFAKA
AHAPFISHPAEFCHLLVALKQRV

Appendix 4: EcBioH-C protein sequence (the hexahistidine tag is in bold)

MNNIWWQTKGQGNVHLVLLHGWGLNAEVWRCIDEELSSHFTLHLVDLPGFGRSRGFGALSLA
DMAEAVLQQAPDKAIWLGWSLGGLVASQIALTHPERVQALVTVASSPCFSARDEWPGIKPDVLA
GFQQQLSDDFQRTVERFLALQTMGTETARQDARALKKTVLALPMPEVDVLNGGLEILKTVDLRQ
PLQNVSMPFLRLYGYLDGLVPRKVVPMLDKLWPHSESYIFAKAAHAPFISHPAEFCHLLVALKQR
V**LEHHHHHH**

Appendix 5: EcBioH-x protein sequence

MNNIWWQTKGQGNVHLVLLHGWGLNAEVWRCIDEELSSHFTLHLVDLPGFGRSRGFGALSLA
DMAEAVLQQAPDKAIWLGWSLGGLVASQIALTHPERVQALVTVASSPCFSARDEWPGIKPDVLA
GFQQQLSDDFQRTVERFLALQTMGTETARQDARALKKTVLALPMPEVDVLNGGLEILKTVDLRQ
PLQNVSMPFLRLYGYLDGLVPRKVVPMLDKLWPHSESYIFAKAAHAPFISHPAEFCHLLVALKQR
V

Appendix 6: EcBioH-NC protein sequence (the hexahistidine tags are in bold)

MGSSHHHHHHSSGLVPRGSMNNIWWQTKGQGNVHLVLLHGWGLNAEVWRCIDEELSSHFTL
HLVDLPGFGRSRGFGALSLADMAEAVLQQAPDKAIWLGWSLGGLVASQIALTHPERVQALVTVA
SSPCFSARDEWPGIKPDVLAGFQQQLSDDFQRTVERFLALQTMGTETARQDARALKKTVLALP
MPEVDVLNGGLEILKTVDLRQPLQNVSMPFLRLYGYLDGLVPRKVVPMLDKLWPHSESYIFAKA
AHAPFISHPAEFCHLLVALKQRV**LEHHHHHH**

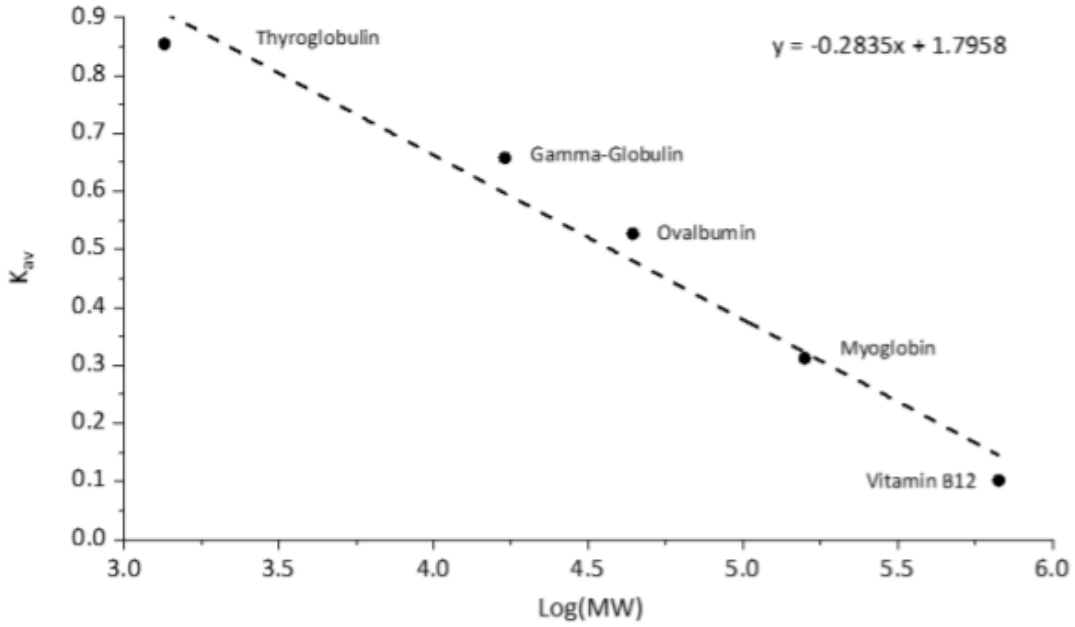
Appendix 7: *H.influenzae bioG* gene sequence

ATGAAAACGAAATTTTACGATTACCAAGGCGAACATTTAATCCTGTATTTTGCAGGTTGG
GGAACGCCGCCCGATGCTGTAAATCATTTGATTTTGCCGGAAAATCACGATTTATTGATT
TGCTATGATTATCAAGATTTAAATTTGGATTTTGATCTTTCCGCCTATCGGCACATCCGT
TTGGTGGCGTGGTCTATGGGCGTTTGGGTGGCAGAGCGAGTATTACAAGGAATAAGATTA
AAATCCGCAACGGCAGTGAATGGCACAGGTTTACCTTGTGATGATAGCTTCGGTATTCCC
TACGCTATTTTTAAAGGTACGCTAGAGAATCTCACAGAAAATACCCGTTTAAAATTTGAA
CGCAGAATCTGTGGCGATAAAGCATCTTTTGAACGTTACCAATTATTTCCAGCCCGTCCG
TTTGACGAAATTCATCAAGAACTTACCGCACTTTTTGCGATGATTCAGCAGGATAAACGC
ATAGATCTTATTCACTGGGCAAATGCATGGGTTAGTTCTCGCGATAAAATTTTTACGCCA
GCTAATCAGCACCAATATTGGGCATTGCGTTGTGCAGTTCAGGAAATAGAGGGTGAGCAT
TATGTGTTTTCAAGATTTACCCACTGGTCGGCATTATGGGATCATTAA

Appendix 8: *H.influenzae* BioG protein sequence

MGSSHHHHHHSSGLVPRGSMKTKFYDYQGEHLILYFAGWGTPPDAVNHLILPENHDLLICYDY
QDLNLDFDLSAYRHIRLVAWSMGVWVAERVLQGIRLKSATAVNGTGLPCDDSFGIPYAIFKGTLE
NLTENTRLKFERRICGDKASFERYQLFPARPFDEIHQELTALFAMIQQDKRIDLIHWANAWVSSR
DKIFTPANQHQYWALRCAVQEIEGEHYVFSRFTHWSALWDH

Appendix 9: Calibration curve of the Superdex HiLoad 16/60 S200 gel filtration column (taken from Dr Piera Marchetti PhD thesis)



Estimation of the protein molecular weight could be determined as followed

$MW(Da) = 10^{a}$

Where $a = (K_{av} - 1.7958) / (-0.2853)$

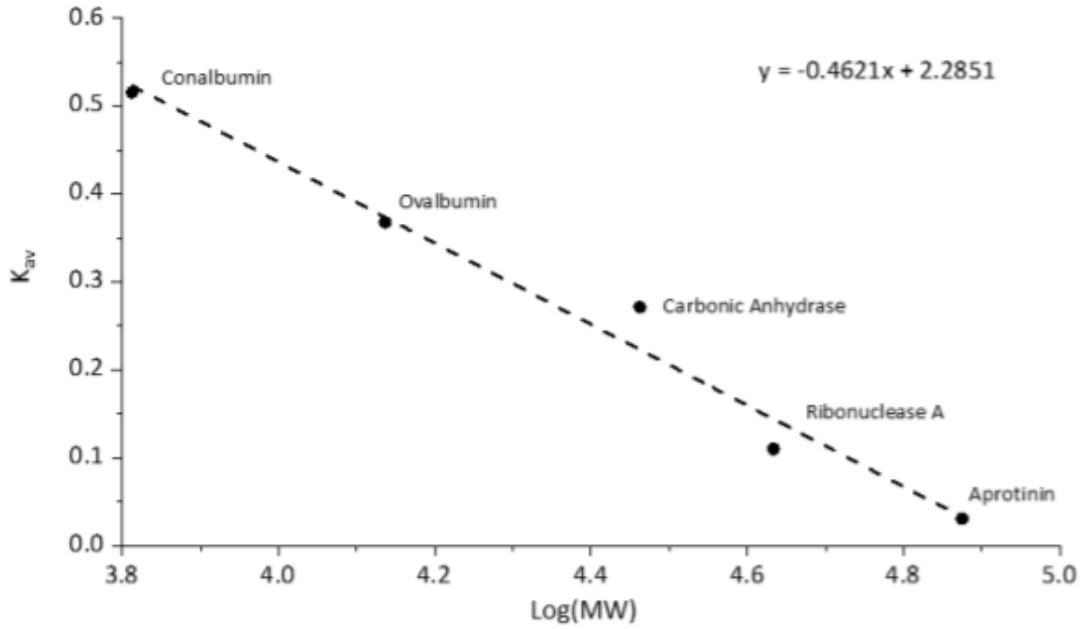
Where $K_{av} = (V_e - V_0) / (V_t - V_0)$

Where V_e = elution volume of the analysed protein

Where V_0 = void volume (41.9 mL)

Where V_t = total bed volume (120 mL)

Appendix 10: Calibration curve of the Superdex HiLoad 16/60 S75 gel filtration column (taken from Dr Piera Marchetti PhD thesis)



Estimation of the protein molecular weight could be determined as followed

$MW(Da) = 10^{a}$

Where $a = (K_{av} - 2.2851) / (-0.4621)$

Where $K_{av} = (V_e - V_0) / (V_t - V_0)$

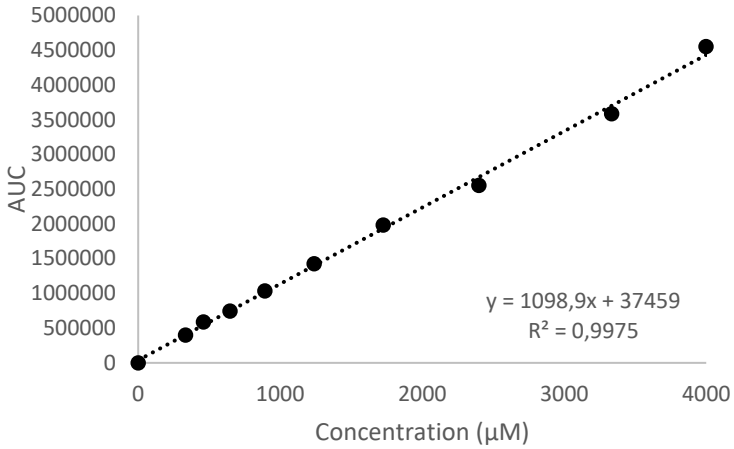
Where V_e = elution volume of the analysed protein

Where V_0 = void volume (41.9 mL)

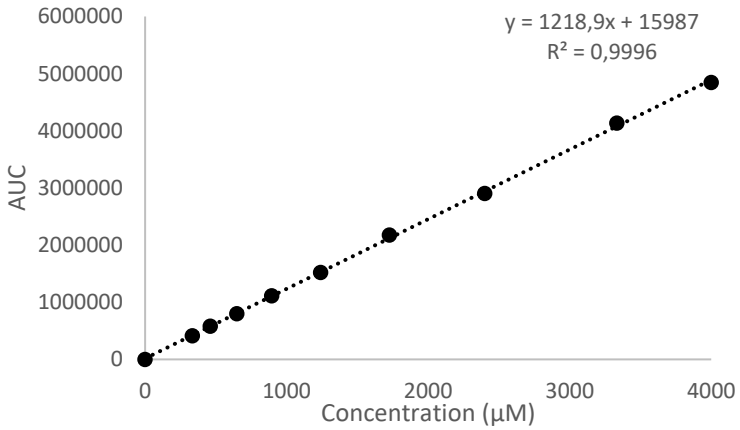
Where V_t = total bed volume (120 mL)

Appendix 11: HPLC calibration curves for 4-NBA and the MBH product

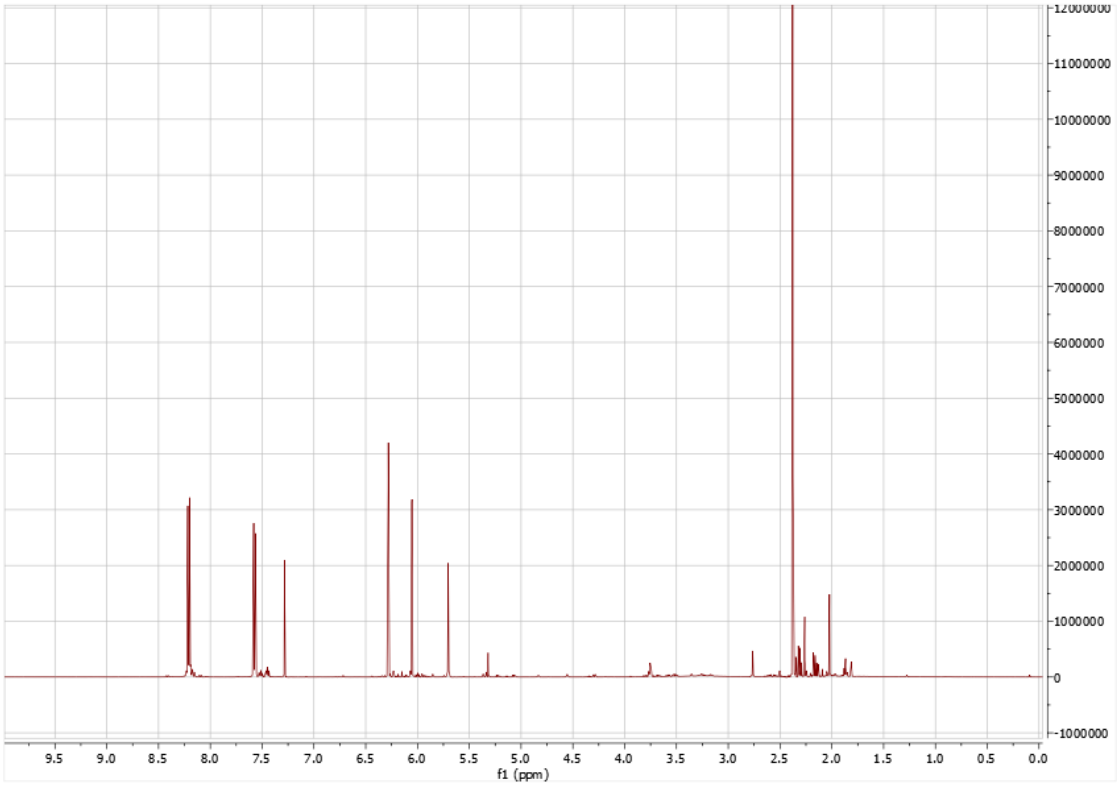
4-nitrobenzaldehyde



MBH product



Appendix 12: ^{1}H NMR of the MBH product



Appendix 13: *B. subtilis* QueC gene sequence

ATGAAAAAAGAAAAAGCAATTGTCGTATTTAGCGGCGGTCAAGACAGCACAACATGCTTA
CTGTGGGCCTTAAAAGAATTCGAAGAAGTCGAAACGGTGACTTTTCATTATAATCAGCGA
CATTCGCAGGAAGTTGAAGTGGCAAAATCCATTGCGGAAAAGCTTGGTGTGAAAAATCAT
TTGCTTGATATGTCACTTTTAAATCAGCTTGCACCGAATGCCCTGACTAGAAATGATATT
GAGATAGAAGTAAAAGACGGCGAACTGCCATCCACATTCGTTCCAGGCCGCAATTTGGTA
TTCTTATCCTTTGCGTCTATCCTGGCTTACCAAATTGGCGCGCGCCACATTATTACAGGA
GTTTGCGAGACAGACTTCAGTGGTTATCCTGACTGCCGTGACGAATTCGTGAAATCTTGC
AATGTCACGGTAAATCTGGCAATGGAGAAGCCGTTTGTGATCCACACGCCTCTCATGTGG
CTCAATAAGGCGGAAACGTGGAAGCTTGCAGATGAGCTTGGCGCGCTGGATTTTGTGAAA
AACAACACGCTGACATGCTATAACGGCATCATCGCAGACGGCTGCGGCGAATGTCCGGCA
TGTCACCTTCGTTCAAAAGGTTATGAAGAATATATGGTGATGAAAGGAGAGCGTGCATAA

Appendix 14: BsQueC protein sequence (the hexahistidine tag is in bold)

MGSSHHHHHHSSGLVPRGSMKKEKAIVVFSGGQDSTTCLLWALKEFEEVETVTFHYNQRHSQ
EVEVAKSIAEKLGVKNHLLDMSLLNQLAPNALTRNDIEIEVKDGELPSTFVPGRNLVFLSFASILAY
QIGARHIITGVCETDFSGYPDCRDEFVKSCNVTVNLAMEKPFVIHTPLMWLNKAETWKLADELG
ALDFVKNNTLTCYNGIIADGCGECPACHLRSKGYEEYMVMKGERA

Appendix 15: *B. subtilis* QueF gene sequence

ATGACGACAAGAAAAGAATCAGAATTAGAAGGTGTAACATTGCTAGGCAATCAAGGTACA
AATTATTTGTTCGAATATGCACCGGACGTGCTGGAATCCTTCCCTAATAAACATGTAAAC
CGTGATTACTTTGTAAAATTCAATTGCCCGGAATTCACATCTTTATGTCCTAAAACAGGC
CAGCCTGACTTTGCGACAATCTACATCAGCTACATTCCTGATGAAAAAATGGTTGAAAGC
AAATCATTAAAGCTGTATCTATTCAGCTTCAGAAACCATGGTGACTTCCACGAGGACTGC
ATGAATATCATCATGAACGACTTGATTGAATTAATGGACCCGCGCTACATTGAAGTATGG
GGCAAATTCACGCCAAGAGGCGGAATTTCCATTGATCCGTACACAAACTACGGAAAGCCT
GGCACGAAGTATGAGAAAATGGCCGAATACCGTATGATGAACCATGATTTGTATCCGGAG
ACAATTGATAATCGTTAA

Appendix 16: BsQueF protein sequence (the hexahistidine tag is in bold)

MTTRKESELEGVTLLGNQGTNYLFEYAPDVLESFPNKHVNRDYFVKFNCPEFTSLCPKTG
QPDFATIYISYIPDEKMVESKSLKLYLFSFRNHGDFHEDCMNIIMNDLIELMDPRYIEVW
GKFTPRGGISIDPYTNYGKPGTKYEKMAEYRMMNHDLYPETIDNR**LEHHHHHH**

Appendix 17: *S. rimosus* ToyM gene sequence

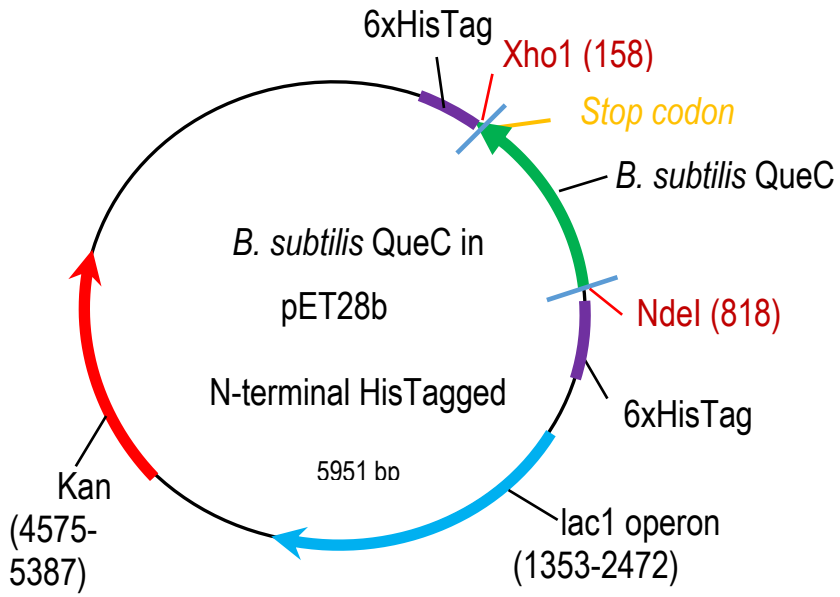
TTGCCCGAATCGCCGGTCGAGGCCGACCCGTTCAGCGCCGTCGTCGTGCTGTCCGGCGG
CATGGACTCCACGACGCTGCTCGCCCATTACGCCATGCTGCGCTACCGGCTCACCGCGGT
GACGGTCGACTACGGTCAGCGCCACCGCCGGGAGATCGACGCGGCCCGGACGATCGCCG
GGCACTACGGCGCGGAACACCATGTGGTGGACCTGAGCGGTATCGGGGTCCTGCTGGGC
GGCTCCGCCCTGACCGACAGTGAGGTCGACGTCCCCTCAGGGCACTACGCCGAGGAGTC
GATGCGGGCCACCGTCGTCCCCAACCGCAATGCCATCCTGGCCAATGTGGCGGTGGGCGC
CGCGGTCGCGCGGCAGGCCGGCATCGTGGCTCTCGGAATGCACGCCGGTGACCATTTCGT
CTATCCCGACTGCCGGCCGGCGTTCGTCGAGGCGCTGGACGGACTCGTCCGGGTCGCGA
ACGAGGGATTCGCCACGCCGCGGGTGGAAGCGCCATTCATCCACTGGAGCAAGACGGACA
TCGCGCGGTACGGGAGCCGGCTGGAAGCCCCGCTGGACAAGAGCTGGTCGTGCTACCGC
GGCGGAGAACTCCACTGCGGTACCTGCGGTACGTGTTACGAACGCCGTGAGGCATTCCGT
GACGCGGACGTGCCGGACCCGACGGAATACCTCGACGCCGCCACCGAATTCGCAGCGCC
GTGA

Appendix 18: SrToyM protein sequence (the hexahistidine tag is in bold)

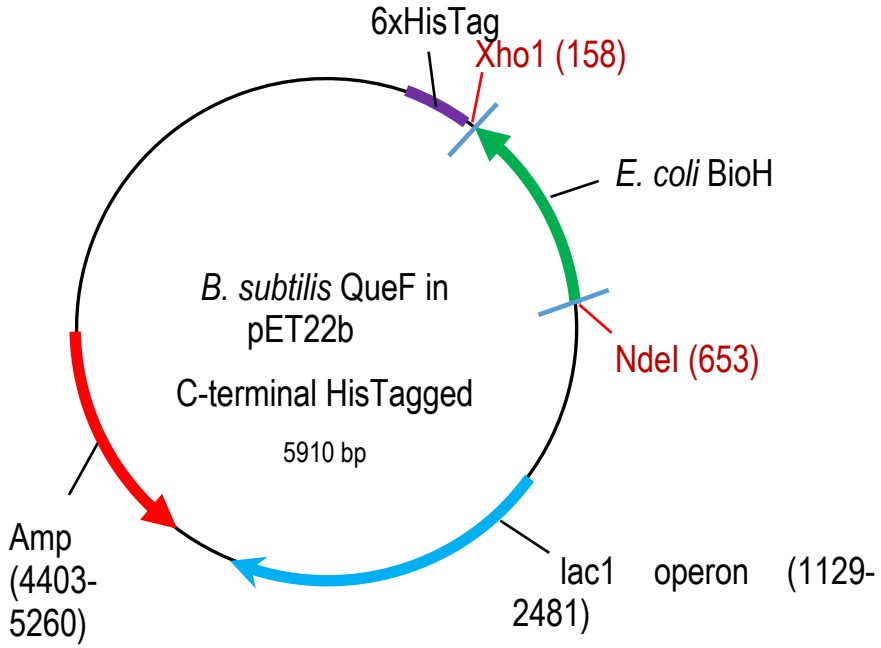
MGSSHHHHHHSSGLVPRGSMPESPVEADPFSAVVVLSGGMDSTTLLAHYAMLRYRLTAVTVD
YGQRHRREIDAARTIAGHYGAEHHVVDLSGIGVLLGGSALTDSEVDVPSGHYAEESMRATVVPN
RNAILANVAVGAAVARQAGIVALGMHAGDHFVYPDCRPAFVEALDGLVRVANEGFATPRVEAPF
IHWSKTDIARYGSRLEAPLDKSWSCYRGGELHCGTCGTCYERREAFRDADVPDPTEYLDAATE
FAAP

Appendix 19: NS vector maps

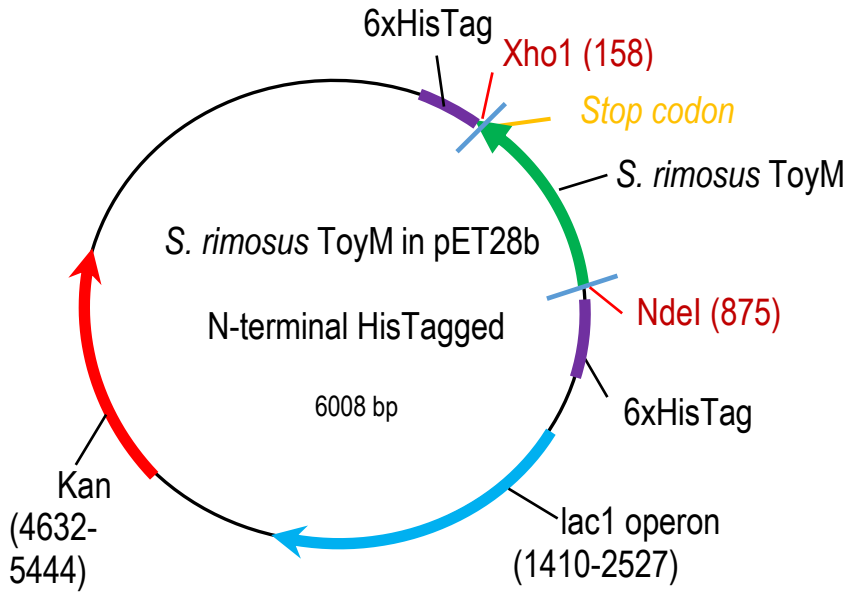
BsQueC



BsQueF

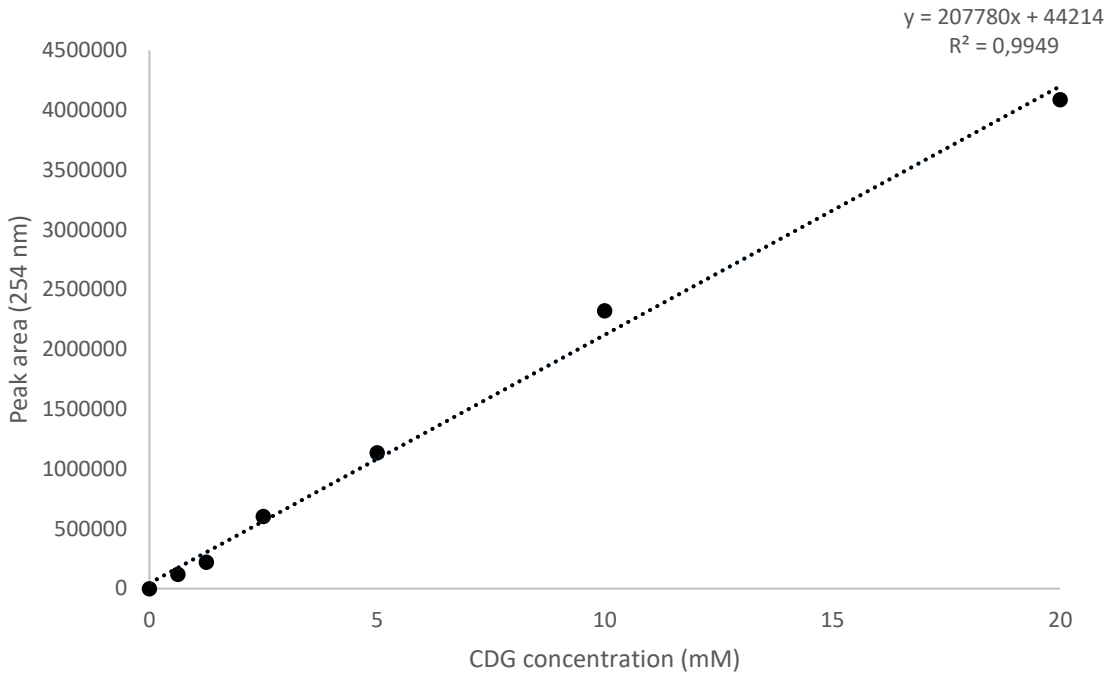


SrToyM

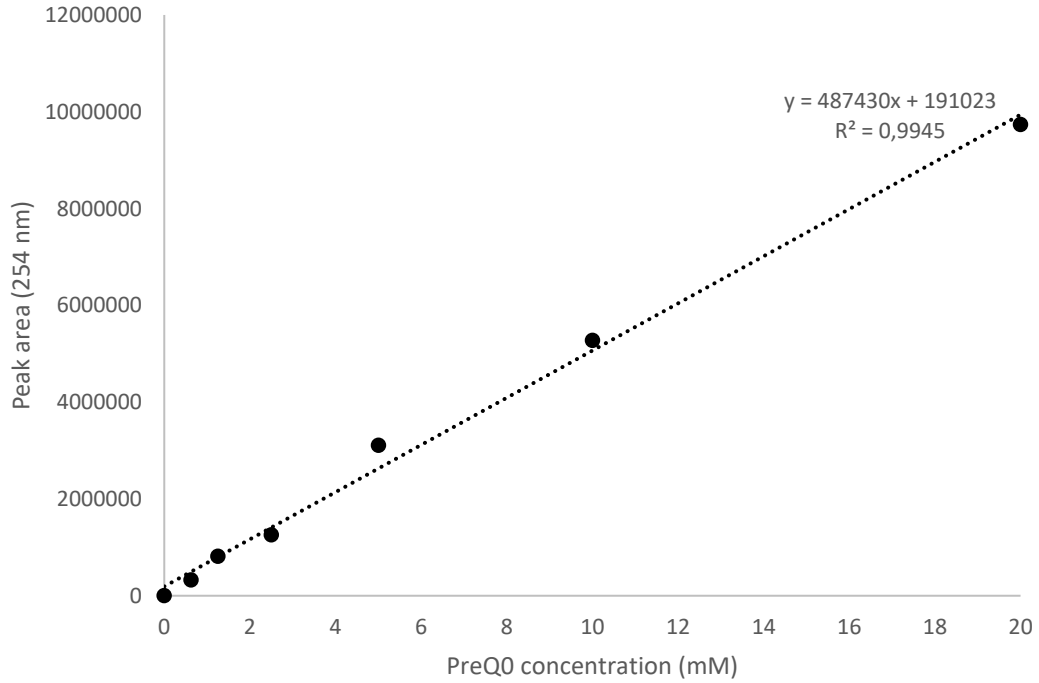

6xHisTag
Xho1 (158)
Stop codon
S. rimosus ToyM
S. rimosus ToyM in pET28b
N-terminal HisTagged
6008 bp
NdeI (875)
6xHisTag
Kan
(4632-
5444)
lac1 operon
(1410-2527)

Appendix 20: HPLC calibration curves for CDG and $PreQ_0$

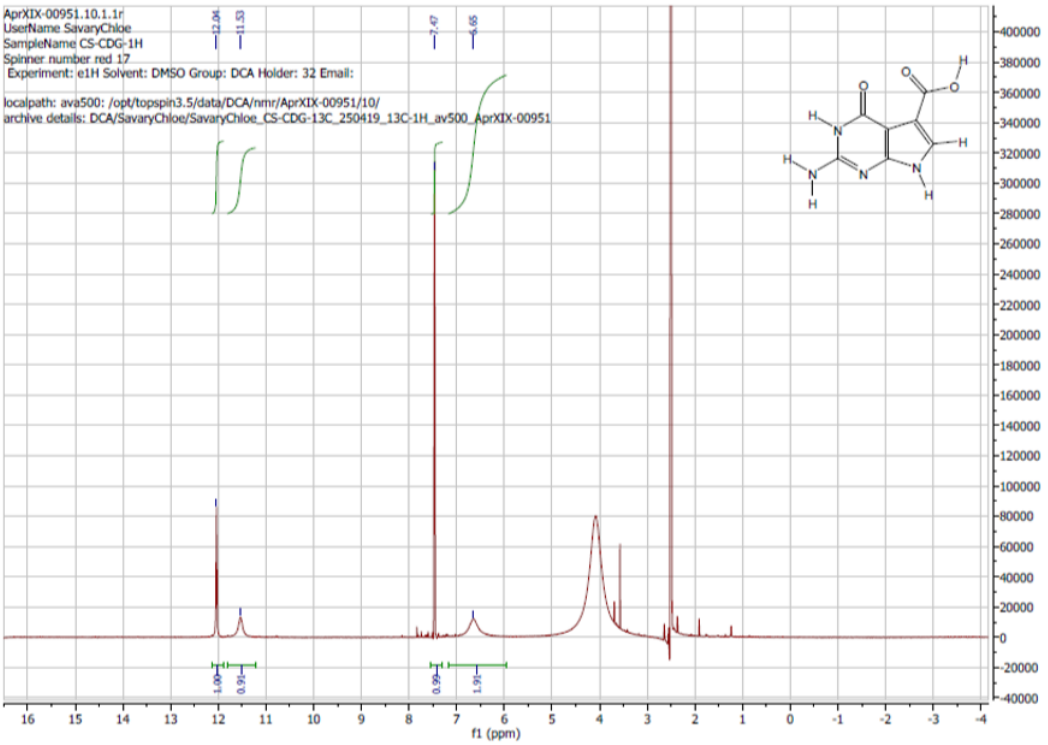
7-carboxydeazaguanine (CDG)



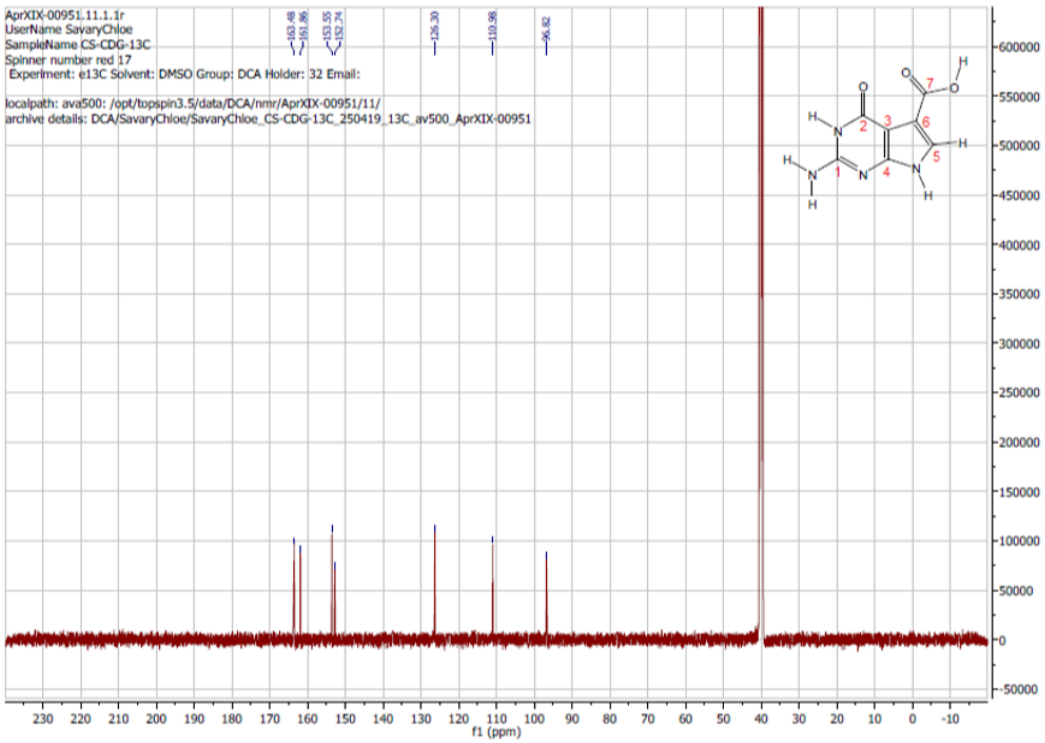
7-cyanodeazaguanine ($PreQ_0$)



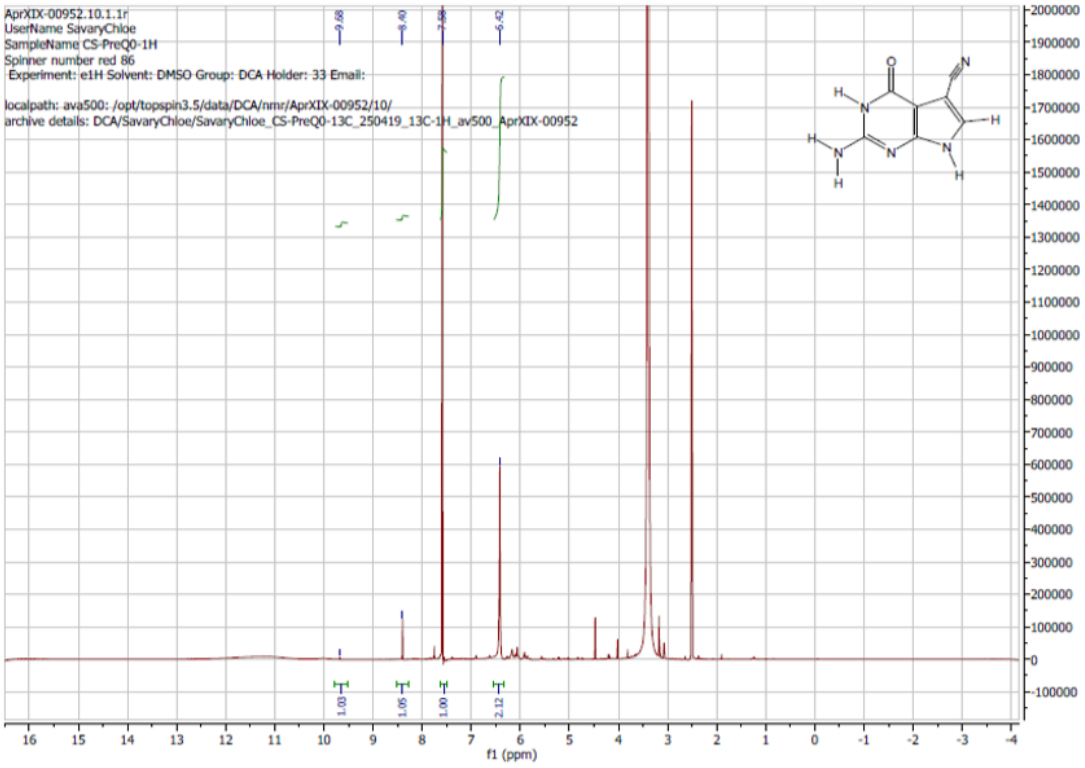
Appendix 21: 7-carboxydeazaguanine ^{1}H NMR



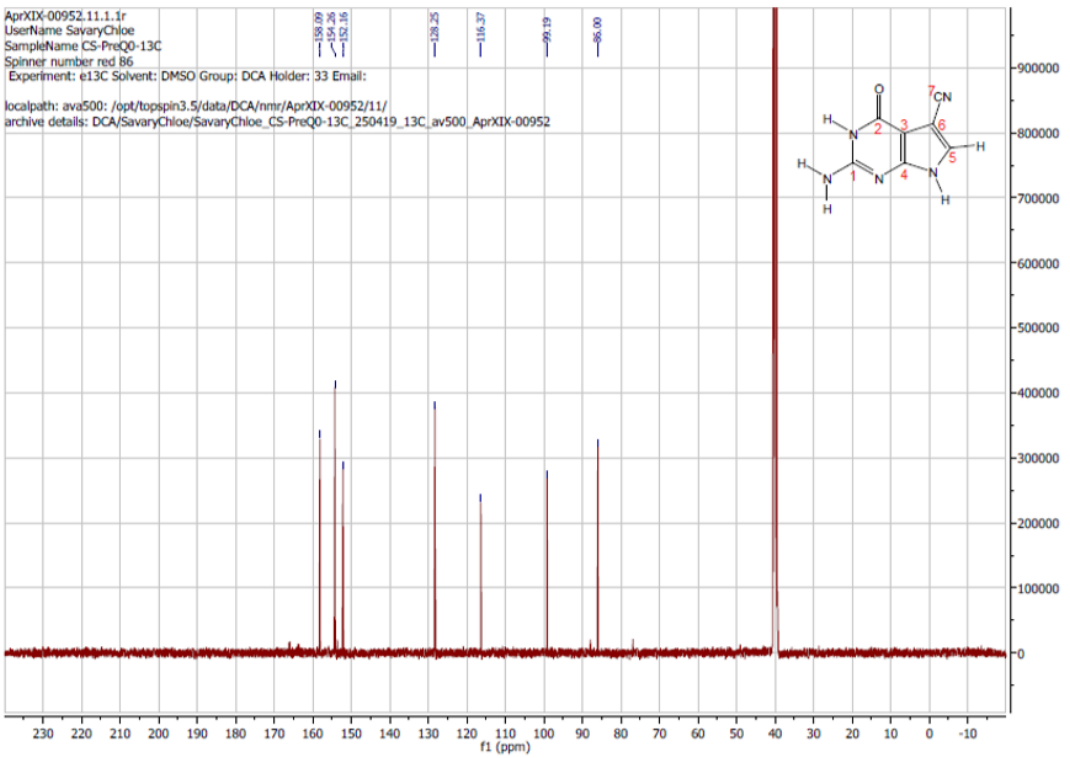
Appendix 22: 7-carboxydeazaguanine ^{13}C NMR



Appendix 23: 7-cyanodeazaguanine ^{1}H NMR



Appendix 24: 7-cyanodeazaguanine ^{13}C NMR



Appendix 25: BsQueC crystallography data refinement and refinement statistics

Wavelength	0.9762	**Ramachandran allowed (%)**	7.78
Resolution range	48.6 - 2.01 (3.118 - 2.01)	**Ramachandran outliers (%)**	1.41
Space group	C 1 2 1	**Rotamer outliers (%)**	0.20
Unit cell	173.477 74.584 135.076 90 93.211 90	**Clashscore**	18.46
Total reflections	127413 (12020)	**Average B-factor**	154.95
Unique reflections	34216 (3260)	**macromolecules**	154.95
Multiplicity	3.7 (3.7)		
Completeness (%)	98.76 (94.98)		
Mean I/sigma(I)	8.15 (0.89)		
Wilson B-factor	93.53		
R-merge	0.09564 (1.467)		
R-meas	0.1117 (1.713)		
R-pim	0.05699 (0.8734)		
CC1/2	0.998 (0.514)		
CC*	1 (0.824)		
Reflections used in refinement	34078 (3235)		
Reflections used for R-free	1648 (204)		
R-work	0.2827 (0.5349)		
R-free	0.3612 (0.5986)		
CC(work)	0.914 (0.699)		
CC(free)	0.910 (0.522)		
Number of non-hydrogen atoms	9129		
macromolecules	9129		